2010

Automated Writing Evaluation for non-native speaker English academic writing: The case of IADE and its formative feedback

Elena Cotos
Iowa State University

Automated Writing Evaluation for non-native speaker English academic writing: The case of IADE and its formative feedback

by

Elena Cotos

A dissertation submitted to the graduate faculty

In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Applied Linguistics and Technology

Program of Study Committee:
Carol Chapelle, Co-Major Professor
Dan Douglas, Co-Major Professor
Nick Pendar
David Russell
Katherine Richardson Bruna

Iowa State University
Ames, Iowa
2010

To my husband Aurel
for his loving support, understanding, encouragement, and patience
throughout the completion of my degree.

To my beautiful daughter Delia
for all the joy she brings to my life
and for making me a happy and proud parent.

To my father Vladimir and my late mother Vera
for their absolute love and for always believing in me.
Mom, you wanted me to live your dream – I am now.

From the bottom of my heart,
Thank you!

Table of Contents

List of Tables

List of Figures

Acknowledgments

This has been a challenging, but ultimately a very rewarding journey, and I am immeasurably indebted to everyone who was part of it. First and foremost, I would like to express my gratitude to my Major Professors, Dr. Carol Chapelle and Dr. Dan Douglas, for their excellent guidance and tremendous support throughout my coursework, research, and the writing of this dissertation. I am honored to have had them as my advisers. My gratitude extends to my committee members for their high expectations and for their contributions at different stages of my dissertation. Thank you to Dr. Nick Pendar, to whom I am obliged for helping me realize my ambitious ideas; without his support and expert contribution, IADE would have not existed. Thank you to Dr. David Russell, who greatly motivated me by believing that my "vision" will take me far in my academic pursuit. Thank you to Dr. Katherine Richardson Bruna, who showed genuine enthusiasm and excitement for what I do every time I met with her. I would also like to gratefully acknowledge the TESL/ALT faculty whose classes I took. Thank you to Dr. Viviana Cortes who showed me the great potential of corpora and who helped me find ways to apply it. Thank you to Dr. John Levis and Dr. Volker Hegelheimer who helped me discover the world of academia and find myself in it. I am also very grateful to Andrew Vernon for his work on the development, integration, and display of IADE's system components on the web.

Thank you to my family and especially to my brother, Eduard Baidaus, who, with his continuous drive and numerous academic achievements, has set a worthy example for me to follow. I would also like to thank John and Linda Hagge for their kindness and gracious hospitality when my family and I first arrived in Ames. Finally, sincere thanks to all my

friends, peer students, and colleagues in the Graduate College for their support and encouragement.

The research presented in this dissertation was supported by the Grant for Doctoral Research in Second/Foreign Language Assessment awarded by Educational Testing Service.

Abstract

This dissertation presents an innovative approach to the development and empirical evaluation of Automated Writing Evaluation (AWE) technology used for teaching and learning. It introduces IADE (Intelligent Academic Discourse Evaluator), a new web-based AWE program that analyzes research article Introduction sections and generates immediate, individualized, discipline-specific feedback. The major purpose of the dissertation was to implement IADE as a formative assessment tool complementing L2 graduate-level academic writing instruction and to investigate the effectiveness and appropriateness of its automated evaluation and feedback. To achieve this goal, the study sought evidence of IADE's Language Learning Potential, Meaning Focus, Learner Fit, and Impact qualities outlined in Chapelle's (2001) CALL evaluation conceptual framework.

A mixed-methods approach with a concurrent transformative strategy was employed. Quantitative data consisted of Likert-scale, yes/no, and open-ended survey responses; automated and human scores for first and last drafts; pre-/post test scores; and frequency counts for draft submission and for access to IADE's Help Options. Qualitative data contained students' first and last drafts as well as transcripts of think-aloud protocols and Camtasia computer screen recordings, observations, and semi-structured interviews.

The findings indicate that IADE can be considered an effective formative assessment tool suitable for implementation in the targeted instructional context. Its effectiveness was a result of combined strengths of its Language Learning Potential, Meaning Focus, Learner Fit, and Impact qualities, which were all enhanced by the program's automated feedback. The strength of Language Learning Potential was supported by evidence of noticing of and focus

on discourse form, improved rhetorical quality of writing, increased learning gains, and relative helpfulness of practice and modified interaction. Learners' focus on the functional meaning of discourse and construction of such meaning served as evidence of strong Meaning Focus. IADE's automated feedback characteristics and Help Options were appropriate for targeted learners, which speaks of adequate Learner Fit. Finally, despite some negative effects caused by IADE's numerical feedback, overall Impact, exerted at affective, intrinsic, pragmatic, and cognitive levels, was found to be positive due to the color-coded type of feedback.

The results of this study provide valuable empirical knowledge to the areas of L2 academic writing, formative assessment, AWE, and I/CALL. They have important practical and theoretical implications and are informative for future research as well as for the design and application of new learning technologies.

Chapter 1. Introduction

1.1. Statement of the problem

The academic writing skills award graduate students admission to their disciplinary discourse communities and potential academic success; therefore, learning to write academically is crucial for international students in higher education in English-speaking countries. Novice academic writers need to learn to produce texts that not only follow general standards of good writing, but also meet the specific demands of their disciplines. Hyland (2000) argues that learning to write academically entails becoming an individual competent in the ways of the given academic community.

> … [I]t must also involve them in acquiring a metacognitive awareness of these forms and contexts and a familiarity with the discoursal strategies they need to perform roles, engage in interactions, and accomplish goals in the target community. In sum, it requires that students gain an awareness of the discipline's symbolic resources for getting things done by routinely connecting purposes with features of texts. (Hyland, 2000, p. 145)

Helping non-native speakers (NNS) become successful academic writers has been the subject of considerable discussion and investigation, the major challenge here being how to cope with discipline specificity. From a number of pedagogical methods, corpus-based approaches have been widely preferred by many EAP practitioners because they provide students with computerized tools for identifying the linguistic features of registers and genres (Aston, 2002; Bernardini, 2000, 2002; Cheng, Warren, & Xu, 2003; Flowerdew, 1998;

Ghadessy, Henry, & Roseberry, 2001; Hyland, 2000, 2003; Kettemann & Marko, 2002; Thompson & Tribble, 2001). Indeed, the use of specialized corpora is a "powerful methodology-technology" for "determining how disciplines use language in their major genres" (Lee & Swales, 2006, p. 57); however, students' problems go beyond linguistic appropriateness. Significant difficulties have been identified at the rhetorical level as well (Cooley & Lewkowicz, 1995, 1997; Dong, 1998; Jenkins et al., 1993; Parry, 1998; Thompson, 1999).

1.1.1. NNS academic writing

Although graduate NNS writers generally have a relatively high level of language proficiency, they still encounter various difficulties when writing for academia. These difficulties have been identified and analyzed in terms of student or supervisor perceptions. Student surveys revealed that graduate NNS students still struggle with vocabulary choice, grammar, and mechanics (Casanave & Hubbard, 1992; Dong, 1998) as well as with organizing ideas and building arguments (Cooleey & Lewkowicz, 1995, 1997). Moreover, they find it very challenging to present their interpretations in ways that meet the expectations of the targeted discourse community (Riazi, 1997). Supervisors, in turn, point to their students' problems at sentence and paragraph levels. Like the students, professors are very concerned about incoherence and faulty development of arguments, finding these difficulties more troubling than faulty grammar (Cooleey & Lewkowicz, 1995, 1997). These are justifiable concerns since even the writing of postgraduate NNS students was found to have such problems as sequencing, developing propositions, and using correct transitions to provide coherence (Dong, 1998).

NNS' academic writing difficulties have also been studied in terms of their impact on overall communicative effectiveness. In a case study of a NNS PhD student, James (1984) found that the sentence-level problems were distracting to the reader, while ineffective organization, lack of coherence, and inappropriate weighing of propositions considerably affected the understanding of the intended meaning. Moreover, he argued that such deficiencies as unnecessarily complex sentences, inadequate referencing, and inappropriate signposting can cause a breakdown of meaning. At the same time, the meaning can also be obscured by unsuitable lexical choices (James, 1984; Cooleey, & Lewkowicz, 1995, 1997).

Genre awareness is a more challenging problem highlighted by previous research. A number of publications emphasize NNS writers' difficulties related to the organization typical of the research reports genre (Cooley & Lewkowicz, 1995, 1997; Dong, 1998; Parry, 1998). In other words, the problem is that NNS students do not adhere to the structural and content conventions specific for individual parts of this particular type of extended discourse. Other studies on thesis writing (Cadman, 1997; Frost, 1999; Knight, 1999) discovered that a major concern was students' ability to critically evaluate theories or methodologies and to develop and support counterarguments with relevant evidence. Similarly, "structuring the text so that the literature contextualizes and illuminates the problem" (San Miguel & Nelson, 2007) was one of the challenges faced by practice-based doctorates. In addition, NNS writers are prone to over- or underestimate their findings, especially when discussing their research results in comparison with those in published literature (Cooley & Lewkowicz, 1995, 1997; Parry, 1998).

Thus, research has demonstrated that NNS graduate students have complex writing

difficulties, the most vivid ones being caused by inappropriate rhetorical and structural choices. Although professors seem to be lenient towards grammatical and vocabulary surface-level errors in NNS' papers, they are not tolerant of "global errors" in their students' writing (Braine, 2002, p. 65). In fact, Jenkins et al. (1993) found that NNS students were evaluated by the same standards as the NS students. Therefore, not mastering discipline-specific academic discourse development norms in English is very likely to affect NNS' academic success. Braine (2002) claims that "NNS students may need explicit instruction in academic writing" (p. 65).

1.1.2. Academic writing pedagogy

Academic writing courses for graduate NNS students should both provide them with general writing practice and "help initiate writers into their field-specific research communities" (Frodesen, 1995, p.333). In support of this opinion, Kushner (1997) also argues that graduate ESL courses have to combine language and discourse with the skill of writing within professional norms. Typically, universities in English-speaking countries offer academic writing courses that bring together students from various graduate programs, e.g., Chemistry, Physics and Astronomy, Architecture, Computer Engineering, Curriculum and Instruction, Molecular Biology, etc., and such heterogeneity makes the instruction very challenging. Nevertheless, there have been a number of attempts to address the writing needs of NNS students in such courses.

At Iowa State University, the NNS graduate-level academic writing course has undergone different changes to bridge the gap between language instruction and disciplinary writing norms. It was once a course that included a wide range of assignments, such as

memos, letters, curriculum vitae, description of processes or mechanisms, literature reviews, research reports, critiques, and oral presentations. The number of assignments has gradually decreased, the major focus being placed on the research report genre. Vann and Myers (2001) employed and described an inductive approach to the analysis of research writing in specific disciplines, in which students examined the format, content, and grammatical and rhetorical conventions of each section of the research report. Supplements to this approach were tasks that required students to write journal entries about the rhetorical conventions of prominent journals in their disciplines and tasks that placed the writing-up research experience "in the framework of an interactive, cooperative effort with cross-cultural interaction" (Vann, & Myers, 2001, p. 82).

Later, after having followed a primarily skill-based approach, in which students wrote field-specific literature reviews, summaries, paraphrasing, data commentaries, and other discipline-specific texts, Levis and Muller-Levis (2003) reported on transforming the course into a project-based one. The project consisted of carrying out original research, the topic of which, for the purpose of coping with discipline diversity, was the same for all students and was determined by the instructor. From the start, the students were provided with a limited set of articles on, for instance, cross-cultural adjustment, with which they worked to identify potential research questions for further investigation and to write the literature review. This approach placed a heavy emphasis on collaboration as students worked in small groups on developing data-collection instruments and on analyzing data. Oral presentations about the group-research project wrapped up the course.

Although no empirical evaluations were carried out, experience showed that there

were both advantages and disadvantages in these approaches. Advantageous were tasks, collaborative and individual, that helped students become acquainted with academic genres through analyses of authentic texts and that geared them towards the production of writing samples similar to those studied in class. What appeared to be a major drawback was the fact that the above-mentioned approaches did not manage to cope with the problem of disciplinarity. Students in different disciplines worked with a limited number of texts, typically on general topics, examining generic features of a given genre, but not being able to fully discover the peculiarities of academic writing in their fields. For that, students needed to be exposed to a large amount of authentic texts that would be exhibiting the various features of the genre as established by specific research communities, which is why the search for an appropriate pedagogical approach to the teaching of academic writing has focused on the use of corpora.

At present, the academic writing course in question is corpus-based and genre-based, combining a top-down approach to genre analysis and a bottom-up approach to the analysis of corpora (Cortes, 2007). Cortes (2007) explains that it was designed to better address the issues of genre-specificity and disciplinarity. In this version of the course, each student works with a corpus of research articles published in top journals of his/her discipline. Students conduct class analyses of their corpus according to guidelines based on empirical findings in Applied Linguistics about the discourse tendencies in research article writing. Their task is to discover organizational and linguistic patterns characteristic of their particular discipline, report on their observations, and apply the knowledge they gain from the corpus analyses when writing a research article for the final project of the course. This instructional method employs the use of concordancers; however, they are not powerful enough as to assist

students in the development of the rhetorical awareness of certain genres, much less in the development of their skill of achieving rhetorical purposes in writing.

It would be ideal to have field-specific writing classes (Cargill et al., 2001; Hyland, 2002), which would be designed for graduate students in certain disciplines, or to have instructors who would be members of students' discourse communities and who would also be trained in language matters (Levis & Muller-Levis, 2003). In reality, however, practicality issues interfere with the implementation of such courses. Although each of the pedagogical approaches described above has its advantages, they all fail to provide NNS students with sufficient practice and guidance through extensive individualized feedback during the process of learning how to produce more complex academic genres in their fields. For that, the students would need a more interactive technology-enhanced approach, by which they would have sufficient opportunities to practice conveying genre-specific messages in writing and to observe their discourse strengths and weaknesses through computer-generated feedback. In other words, they would need an automated writing evaluation (AWE) system, which would both draw from relevant theoretical models and be responsive to practical needs by analyzing students' texts and providing individual and discipline-specific feedback.

While AWE systems have recently gained much popularity, it is not clear how they can best be implemented in effective pedagogical practice for a number of reasons. First, AWE was initially developed for summative assessment rather than for purposes of remediation, and knowledge of the effectiveness of automated evaluation for formative assessment is scarce. Second, non-native speakers were generally not the targeted learners; previous research findings with native speakers do not necessarily generalize to ESL/EFL

populations. And third, a problem directly related to graduate NNS academic writing is that, to date, there is no AWE program that would be able to analyze academic genres other than the essay and to recognize more sophisticated and subtle rhetorical shifts than thesis and topic sentences. This dissertation aims at addressing these issues in the context of the corpus and genre-based academic writing course offered to NNS graduate students at Iowa State University.

1.2. Dissertation goals

In view of the lack of suitable AWE tools, a preliminary goal for this dissertation was to develop a new program, called IADE (Intelligent Academic Discourse Evaluator), with the ability to evaluate students' research article Introduction sections and to generate intelligent individualized feedback on each submitted draft. The major purpose of the dissertation was to implement IADE as a formative assessment tool that complements academic writing classroom instruction and to investigate the effectiveness and appropriateness of its automated evaluation and feedback.

1.3. Significance of the study

The knowledge acquired in this study makes significant contributions to four different sub-fields of Applied Linguistics - NNS academic writing pedagogy, formative assessment, AWE, and Intelligent Computer Assisted Language Learning (ICALL). The first benefits from an innovative technological approach to instruction, which caters to the needs of both students and their teachers. Second, this study adds to the limited body of knowledge on whether and how non-native speakers benefit from intelligent formative feedback. Third,

given that research on AWE for formative assessment is not plentiful, this work contributes to strengthening the connection between the two areas by demonstrating how they can benefit from each other. Specifically, formative assessment is enhanced by automatization and individualization of feedback, while AWE finds valid and legitimate uses for its technological potential in learning contexts. In other words, this dissertation also expands the focus of AWE validation research by investigating the application of IADE in a specific L2 instructional environment. Finally, this complex work, being grounded in theory and responsive to practical needs, serves as a model for I/CALL development and evaluation research.

1.4. Outline of the dissertation

This first chapter introduced a number of problems in graduate-level NNS academic writing instruction, explaining how those motivated the research goals of this dissertation. Chapter 2 reviews the literature on topics relevant to the current study, providing a theoretical framework that draws on Systemic Functional Linguistics (SFL), Interactionist views on Second Language Acquisition (SLA), and the Skill Acquisition Theory (SAT). The purposes of formative assessment and the features of computerized formative feedback, as well as a discussion of AWE systems and their uses are also included in this chapter. Chapter 3 elaborates on the context of the study and on the instructional tools involved, describing the design of IADE from theoretical, empirical, and development perspectives. Chapter 4 presents the research approach and the CALL conceptual framework (Chapelle, 2009) that guides this study and provides a detailed description of the participants, materials, procedures, and data analysis. The results on Language Learning Potential, which is a central

concept in the guiding conceptual framework and therefore the most substantial part of this study, are elucidated in Chapter 5. Chapter 6 presents the findings on three other concepts pertaining to the chosen framework - the Meaning Focus, Learner Fit, and Impact qualities of IADE. Finally, Chapter 7 summarizes and evaluates the obtained results, bringing into the discussion the limitations of this study and the multifaceted implications of its findings. It concludes with outlining directions for future research.

Chapter 2. Literature review

This chapter is a literature review that elaborates on three major perspectives undergirding this dissertation research: theoretical, assessment, and technological. The theoretical framework that provides a sound basis for the instructional context of this study includes the Systemic Functional Linguistics (SFL), the Skill Acquisition Theory (SAT), and the Interactionist approach to Second Language Acquisition (SLA). All these have been identified as relevant and informative for L2 academic writing classroom instruction, but, at the same time, it was argued in this chapter that some essential insights from these theories have weak links to current pedagogical practice. First, key SFL concepts such as texts as units of analysis and genre receive considerable attention in the teaching of academic writing, while the emphasis on contextual richness and functional meanings is unsubstantial. Second, the presentation, practice, and production development stages postulated by SAT are viewed as being reflected in the types of instructional tasks completed by students; however, the practice element, which is crucial for skill automatization, is limited to the production of single drafts. And third, interaction and feedback, which interactionists consider to play a significant role in second language acquisition, are also deficient.

Further, the insufficiency of individualized feedback is also problematic in terms of formative assessment. In order to elucidate the nature of this problem, this chapter introduces the purposes of formative assessment and the qualities of formative feedback. Due to the potential of computerized feedback to enhance formative assessment, ICALL systems and the intelligent feedback they generate are reviewed here as well. This review reveals that research on intelligent formative feedback is very scarce. Moreover, it is argued that any type

of such feedback needs to be soundly conceptualized and operationalized, and therefore the Evidence Centered Design (Mislevy et al., 2003) is proposed as a suitable framework.

The third, technological perspective of automated writing evaluation (AWE), serves the purpose of addressing the theoretical and formative assessment issues. To demonstrate how this purpose can be served, AWE systems such as Criterion, MyAccess!, and WriteToLearn are described. Also, the features of these programs are used to highlight the advantages of AWE and to discuss how this technology can enhance the most essential qualities of formative feedback. Given that AWE has not only supporters but also opponents, issues of expressed concern related to the consequences of AWE classroom implementations are pondered over here as well, and a review of the few existing studies is provided.

After reviewing the literature and highlighting major gaps, which provide a solid justification for the evaluative nature of this dissertation work, the evaluation approach is introduced. It is followed by a set of research questions formulated in view of the conceptual framework that was chosen to guide the study.

2.1. Theoretical framework

2.1.1. Systemic Functional Linguistics (SFL)

Systemic-Functional Linguistics (SFL) is a theory of language that considers language function and semantics as being central to communicative activity. In other words, SFL centers on the function, or use, and meaning of language as opposed to the elements of language structure and their combinations that abide by certain rules. Language here is viewed as a systematic resource for meaning called 'meaning potential' (Halliday, 1971),

which is used in a particular context. The context is considered as defining the meanings that can be expressed and the range of language choices that can be used in expressing those meanings.

According to SFL, the most appropriate units for analysis are texts, which are simply defined as language that is functional, where 'functional' means “language that is doing some job in some context, as opposed to isolated words or sentences” (Halliday, 1989a, p.10). From this perspective, texts are made of meanings, which, in order to be expressed, have to be coded in words and structures. Further, SFL maintains that texts are produced because of their environments, i.e. the interactive events of social exchange of meanings, which are continuous processes of choices in meanings that are represented through the linguistic system. The context unfolded in the text is encapsulated “through a systematic relationship between the social environment on the one hand, and the functional organization of language on the other” (Halliday, 1989a, p.11).

The language system is organized into functional components identified as ideational, interpersonal, and textual (Halliday, 2007). The first, ideational, aspect of the organization of the semantic system includes experiential meaning, which reflects a representation of reality as apprehended in one's experience, and logical meaning, which refers to the expression of fundamental logical relations among ideational elements. In its interpersonal function, language is a way of acting in that it is a piece of action between interlocutors. The last, textual meaning, comprises features such as semantic and grammatical balance, thematic structure, information focus, etc., which make the text a text. Halliday (1989a) explains that

these “strands of meaning are all interwoven in the fabric of discourse” and asserts that “every sentence in a text is multifunctional” (p.23).

These functional components of language are systematically related with the abstract components of field (i.e., what is going on), tenor (i.e., who is involved), and mode (i.e., role assigned to language), which characterize the relation of texts to their contexts. The field is expressed through the experiential function, the tenor – through the interpersonal function, and the mode – through the textual function in the semantic system.

Another important concept in SFL is that of a register, which is considered to be a semantic concept referring to meanings that are associated with the field, mode, and tenor of a particular context of situation. In other words, register is “a variety according to use [...] determined by the nature of activity in which language is functioning” (Halliday, 1989b, p. 43). The semantics of registers can differ, and so can their lexico-grammatical features that realize certain meanings. Some registers can even have indexical features in the form of particular words or grammatical signals that indicate a particular register. For example, 'once upon a time' is an index that denotes a tale. The functions of a research article introduction can be indexed by ‘the increasing interest in’, ‘no work has been reported on’, ‘the goals of this investigation were to’, etc. Furthermore, in some registers, the number of possible meanings may be limited, while in others the range of discourse is not so much constrained at one level. All texts are different from one another because every instance of language use is in some sense unique. At the same time, though, as Halliday theorizes, any text is in some way similar to other texts.

Martin (1992) extends the notion of register, defining it as “metafunctionally organized connotative semiotic between language and genre” (p. 502). Viewed as constraining the possible combinations of field, mode, and tenor variables, genre is considered by systemic functionalists a semiotic system underlying register. At an abstract level, it “represents the verbal strategies used to accomplish social purposes of many kinds. These strategies can be thought of in terms of stages through which one moves in order to realize a genre” (Martin, 1985, p. 251). They are known as 'schematic', or 'generalized', structures (Hasan, 1978; Martin, 1985; Ventola, 1982), which represent ways in which texts perform the functions of given genres in given contexts. For example, according to Ventola (1982), the schematic structure of service encounters consists of:

Greeting (exchange of 'hello', 'good morning', etc.)
Attendance allocation (addressing the next customer: 'Next please')
Service bid (offer of service: 'How can I help you?')
Service (statement of needs: 'Yes, I'm looking for...')
Resolution (decision to or not to buy: 'Yes, I'll have the...')
Pay (exchange of payment)
Goods handover (exchange of goods)
Closing (exchange of 'thanks')
Good-buy (exchange of 'Bye')

The sequence of realization of these elements of the schematic structure may vary. Some elements may be obligatory and some optional. Hasan (1989) argues that “the obligatory elements define the genre to which a text belongs” (p. 62). The optional elements, however, are not random; they can possibly occur as predicted by some attribute of the

contextual configuration, i.e., by the values that realize field, tenor, and mode. In other words, the schematic structures are generated by the register. Martin (1985) inverts this formulation to say that the elements of schematic structure determine particular values of field, mode, and tenor and that "the elements themselves are generated by genre networks" (p. 252). Whatever the theoretical perspective, what is important is acknowledging that the values of register and schematic structure are related.

2.1.2. Relevance of SFL to academic writing

According to Halliday (1985), "a theory of language [is] essentially consumer-oriented" and "the value of a theory lies in the use that can be made of it" (p. 7). He explains that "systemic theory is designed not so much to prove things as to do things" (p. 11). This study is using SFL as opposed to doing SFL. That is, SFL here does not serve as the background theory for studying the language in order to understand how it works and what people do with it; rather, relevant SFL concepts are used to inform and improve decisions concerning the targeted academic writing context.

Halliday (1989b) reasons that the only way to learn how to make texts is by making texts just like the only way to learn a language is by using it. He also explicates that familiarity with different genres is crucial, and that such familiarity "does not grow automatically" (p. 68), advising that learners need to be exposed to genres, "particularly those that are actively required in the educational process" (p. 69). As briefly described in Chapter 1, the graduate academic writing courses at ISU base their instruction on the analysis of texts. The current corpus-based approach is intended to provide the necessary exposure to the research article genre, which the students are required to produce in their academic life.

Research articles reflect a certain academic context and are therefore representative of certain academic or professional discourse communities. Non-native English speaking students entering academic disciplines "need a specialized literacy that consists of the ability to use discipline-specific rhetorical and linguistic conventions to serve their purposes as writers" (Berkenkotter, Huckin, & Ackerman, 1991, p. 191). Since the discourse communities that students need to become members of are not generally located in specific physical settings, the most appropriate way for students to learn to write in a way that meets the expectations of their discourse communities is by learning from the discourse that members of those communities use to communicate with each other.

Given this context, and given that in SFL texts are viewed the most appropriate units for analysis because they are semantic units that exhibit internal cohesiveness and contextual consistency (Halliday & Hasan, 1976, p.1), students' focus on texts in corpus-based academic writing courses is plausible. However, this language theory posits that texts are both products of choices made from the meaning potential available in particular contexts and processes of meaning exchanges between the participants (here, writers and readers). Therefore, texts have an interactive nature and represent social events that unfold linguistically (Hasan, 1978, p.229). Considering that, the academic writing instruction in question is both advantageous and faulty. Advantageous is the fact that the students work with professional texts in their own disciplines in order to learn to engage in social interaction between themselves and their academic audiences by producing texts that model on those they are exposed to when working with their discipline-specific corpora. The disadvantage is that the students' work is mostly oriented towards identifying the structural

components of texts, or, in systemic-functionalist terminology, the schematic structures. Halliday (1982) argues that "[e]very structural feature has its origin in the semantics, that is, it has some function in the expression of meaning" (p. 8). While the students may be taught the functions of certain structural features of the research article sections, the writing tasks do not draw their attention to functional meaning, expressing which, according to SFL, is essential to language learning. Consequently, these academic writing courses need to extend their focus on texts to include a focus on language as a systematic resource that the students should learn to use taking into account the context, the text's rhetorical structure, multifunctionality, and the higher-level semiotics that build the discourse.

This research is situated in a class that aims at teaching international graduate students to produce research articles according to the discourse and linguistic conventions in their fields, which they learn about from discipline-specific corpora. Classroom instruction focuses on such components of the research article genre as Introduction, Methods, Results, Discussion, Conclusion, and Abstracts.

In SFL, genres are viewed as "staged goal-oriented social purposes" (Halliday & Martin, 1993, p. 36), and they are analyzed by being divided into stages, each of which contributes to an overall cultural goal. Since, from this perspective, texts can be compared and contrasted to show a generic structure potential, the research article genre is presented to students in this academic writing class as 'formulae' used in their disciplinary cultures, with certain components of the Introduction, Methods, Results, Discussion, Conclusion, and Abstract sections being compulsory or optional.

This dissertation concentrates on one component of the research article studied in the course – the Introduction section, which is a type of text that is very rich in functional meanings. The approach to teaching and learning how to write Introductions is based on Swales' (1981, 1990, 2004) genre analysis work, which has greatly contributed to centralizing the concept of genre in specialized language teaching and in the development of professional communication skills. Following SFL assumptions that cultural expectations and values can be revealed by examining text organization, Swales proposed the so-called Create-a-Research-Space (CARS) model for research article introductions. This model has been successful for both descriptive and pedagogical purposes due to its simplicity and functionality. It relies on the notion of 'move', which is defined as “discoursal or rhetorical unit that performs a coherent communicative function in a written or spoken discourse” (Swales, 2004, p. 228). The moves are seen as flexible in their linguistic realizations, which may vary from single clauses to several sentences. Swales points out that moves are functional rather than formal units.

According to Swales, research article introductions consist of three moves: Move 1 – Establishing a territory, Move 2 – Establishing a niche, and Move 3 – Occupying the niche, which have a noticeable metaphorical coloring, “that of ecological competition for research space in a tightly contested territory” (Swales, 2004, p. 226). Each move contains a number of ‘steps’, and these steps are fulfilling particular functional meanings, which may be obligatory and/or optional depending on the norms adopted by field-specific discourse communities. Figure 2.1 below presents Swales' move schema.

Move 1: Establishing a territory
Step 1 – Claiming centrality *and/or*
Step 2 – Making topic generalization/s *and/or*
Step 3 – Reviewing previous research

Move 2: Establishing a niche
Step 1A – Indicating a gap *or*
Step 1B – Highlighting a problem *or*
Step 1C – Question-raising *or*
Step 1D – Hypothesizing *or*
Step 1E – Adding to what is known *or*
Step 1F – Presenting justification

Move 3: Occupying the niche
Step 1A – Announcing present research descriptively *or*
Step 1B – Announcing present research purposefully
Step 2A – Presenting RQs *and/or*
Step 2B – Presenting hypotheses
Step 3 – Definitional clarifications *and/or*
Step 4 – Summarizing methods *and/or*
Step 5 – Announcing principal outcomes *and/or*
Step 6 – Stating the value of the present research *and/or*
Step 7 – Outlining the structure of the paper

Figure 2.1. Swales' Move schema (based on Swales 1981, 1990, 2004)

Although the moves and their possible steps are so clear-cut, move identification is not as transparent as it may seem. Nwongu (1990) claims that it is more of a bottom-up process, which is at the same time influenced by one's schemata about the structuring of text-type and genres. Swales (2004), however, suggests that, in some cases, certain lexico-grammatical features can indicate certain moves. For instance, the present continuous tense can invoke recency in statements of Centrality in Move 1, lexical units with negative connotations can indicate a gap or a problem, and deictics and personal pronouns can signal the onset of Move 3. In other cases, the placement of a discourse piece can help to interpret its function.

2.1.3. Skill Acquisition Theory (SAT)

Practitioners have increasingly adopted a view on writing as a process rather than a product. The process-based approach emphasizes learning how to organize texts as well as how to effectively express meaning (Matsuda, 2003). Invention strategies, multiple drafts, and formative feedback became essential elements of writing instruction aimed at assisting students in the development of their writing skills. Since skill development is of major concern, relevant to this instructional goal is the Skill Acquisition Theory, which is rooted in Psychology (Newell & Rosenbloom, 1981). DeKeyser (2007) explains that "learning a wide variety of skills shows remarkable similarity in development from initial representation of knowledge through initial changes in behavior to eventual fluent, spontaneous, largely effortless, and highly skilled behavior" (DeKeyser, 2007, p. 97). Byrne (1986) postulated three stages of skill development: presentation, practice, and production. These stages were elsewhere referred to as cognitive, associative, and autonomous (Fitts & Posner, 1967) and declarative, procedural, and automatic (Anderson, 1982). At the first stage, learners acquire some knowledge about a certain skill, without attempting to apply that knowledge in practice. Acting on this knowledge and transforming it into a behavior, or, in DeKeyser's words "turning 'knowledge that' into 'knowledge how'" (2007, p. 98), occurs at the second stage, which does not necessarily require much time if the declarative knowledge is well acquired. Considerable practice is needed, though, before that particular behavior becomes automatic at the third stage. Practice is the key to gradual qualitative change in the cognitive mechanisms activated by learners to execute a given task.

2.1.4. Relevance of SAT to academic writing

How do these theoretical insights apply to the teaching of academic writing? As mentioned before, in genre and corpus-based courses at ISU, the students are taught the basics of the research article genre, which means that they are provided with presentation that equips them with new, declarative knowledge. When studying the writing conventions of research article Introductions, the declarative knowledge is acquired through guided observation and analysis of the discourse moves in the corpora. The problem is that typically the final task in the course requires the students to draw on their declarative knowledge and produce a research article, abiding by the norms in their disciplines, without being provided with sufficient opportunities for targeted writing practice. This means that the students are expected to skip the second stage of the skill development, but still perform well. To address this problem, the academic writing instructional approach in question needs to add a core element – that of practice. Practice is more likely to become helpful and meaningful if it is motivated by interactive feedback according to the Interactionist Approach to SLA.

2.1.5. Interaction approach to Second Language Acquisition (SLA)

Educators working in the field of second language writing see feedback as being crucial for encouraging and consolidating learning. While the importance of feedback was articulated by the social cognitivist perspective in cognitive psychology and emphasized with the emergence of learner-centered approaches to L2 writing, it was also embraced by the SLA interactionist approach, according to which learning occurs "through the learner's exposure to language, production of language, and feedback on that production" (Gass & Mackey, 2007, p. 176). In other words, the Interaction Hypothesis states that such constructs

as input (Krashen, 1982, 1985), interaction (Pica, 1998), feedback, and output (Swain, 1985, 1995) play a significant role in second language acquisition.

This theoretical perspective, as well as other approaches to SLA, recognizes input as an essential component for learning because it is the source for learners' potential linguistic hypotheses. Modified input, which can take the form of simplification or elaboration, makes the language more comprehensible for learners. Interaction, in turn, is the context in which the language is used and in which learners can receive feedback that is either confirming their communicative success or pointing to failure in their production, thus helping them notice problematic language use. The feedback also stimulates testing hypotheses, which learners may generate with regards to the nature of their linguistic problem, and prompts revised output. Continued production of output is important because, in the long run, it leads to automaticity and better fluency. The relationship between these constructs is established by negotiation of meaning, which "triggers interactional adjustments by the NS or more competent interlocutor" and "connects input, internal learner capacities, particularly selective attention, and output in productive ways" (Long, 1996, pp. 451-452).

Traditionally, interaction referred to learners' engagement in conversations with interlocutors. Ellis (1999) expanded the idea of interaction from interpersonal to that of intrapersonal interaction, i.e. "interaction that can occur in our minds, […] and, more covertly, when different modules of the mind interact to construct an understanding of or a response to some phenomena" (p.1). Later, Chapelle (1998; 2003, 2007a) connected these concepts to learner-computer interaction, showing how learning with the help of computers can reflect the interactionist constructs. Specifically, the linguistic features in the input can be

made salient through highlighting, glosses, hyperlinks, pictorial or video representations, etc. The input itself can be automatically adapted or modified through restatements, repetitions, non-verbal prompts, change of presentation modes, and the like.

Various computer-mediated communication applications and automated writing and speech processing software support the provision and generation of immediate feedback. Garret (1987) identifies four types of CALL feedback: (1) only the correct answer is presented; (2) the location of errors on the basis of a letter-by-letter comparison of the learner's input with the machine-stored correct version is indicated; (3) based on an analysis of the anticipated incorrect answers, error messages associated with possible errors are stored in the computer and are presented if the learner's response matches those possible errors; and (4) based on a linguistic analysis of the learner's response compared to an analysis derived from relevant grammar rules and lexicon of the target language, problematic or missing items are pinpointed.

The last type of feedback, known as intelligent feedback, is the most sophisticated in that it uses Natural Language Processing (NLP) approach, which is superior to the so-called 'pattern markup' and 'error-anticipation' techniques used in other conventional types of feedback because it is capable of diagnosing errors, providing detailed explanations about the nature of those errors, and responding to multiple problematic aspects of language use that may occur in learner's production. In addition to feedback, interaction is also enhanced through mouse clicks and hypertext links that provide learners with opportunities to request help on demand, which offers them immediate assistance during the learning process in the form of explanations, examples, definitions, concordancing lines, etc. All these types of

feedback and help options can make learners notice their linguistic infelicities and attempt to produce more comprehensive output. Thus, the SLA interactionist model applied to computer-assisted environments can and should inform and greatly enhance the teaching and learning of academic writing.

2.1.6. Relevance of Interactionist SLA to academic writing

The corpus-based approach employed in L2 graduate academic writing pedagogy draws only partial insights from the SLA Interactionist models. This teaching approach exposes the students to large amounts of input by having them read and analyze the texts and the language in the targeted genre. Conducting corpus analyses is meant to help the students notice the characteristics of the target language. Then, the students are given the opportunity to produce their own written output in the form of research article sections, which have to be modeled on the writing conventions observed in the input corpora. The elements of interaction and feedback, although considered essential in the chain of acquiring language knowledge, are largely in deficit. This prevents the students from noticing infelicities in their output and correcting those as a result of engaging in target language interaction. Interaction is limited to regular in-class teacher-student group instructional, explanatory exchanges and occasional feedback on student papers, but it can be potentially expanded through anytime computerized feedback.

2.2. Formative assessment

2.2.1. Purposes and of formative assessment

As discussed above, a considerable weakness in the teaching of NNS graduate-level academic writing is little targeted writing practice and interaction aimed at providing feedback on student's writing performance, which can be addressed by enhancing the formative assessment aspect of instruction. Formative assessment is distinguished from summative assessment, the definitions of the two revolving around differences in function and timing. Summative assessment measures students' performance at the end of a course or a program with the purpose of grading, certification, or accountability, while formative assessment is conducted during the teaching process for the purpose of identifying students' difficulties and monitoring the instructional process (Brown & Knight, 1994; Ebel & Frisbie, 1991; Torrance & Pryor, 1998). The latter aims at providing students with feedback, which is meant to help them close "the gap between actual and desired levels of performance" (Wiliam & Black, 1996, p.543).

The major purpose of formative assessment provided during instruction is to improve student motivation and learning (McMillan, 2007). Key to this approach is supplying students with directive and facilitative feedback that is intended to modify their thinking or behavior for more successful learning. Formative feedback, to be effective, has to fulfill a number of roles (see Sales, 1993):

- To direct (i.e, to provide information about what and how to improve)
- To inform (i.e, to acknowledge the accuracy of students' production)

- To instruct (i.e, provide additional information to improve students' understanding of what is being learned)
- To motivate (i.e, to provide a motive for continuing effort)
- To stimulate (i.e, to arouse students to continue)
- To advise (i.e, to inform students of the status of their efforts in relation to expected performance)
- To summarize (i.e, to offer a cumulative report of individual performance)

Additionally, formative assessment is believed to be effective when a number of conditions have been met: first, the feedback must be timely; second, it must include "some degree of prescription about what must be done" (Wiliam & Black, 1996, p.543); third, it should not give correct answers; and, fourth, this assessment must provide opportunities for repeating the task upon the reception of feedback (Buchanan, 1998, 2000). Educational psychology clearly outlines the most essential qualities of formative feedback: cognitive involvement, goal-orientation, complexity, specificity, and timing. These qualities will be used in this dissertation in order to conceptualize formative feedback in language learning since such an informative framework is not available in the field of SLA or in writing instruction.

2.2.2. Qualities of formative feedback

- Cognitive involvement

Formative assessment entails teachers' valid and reliable judgments on students' work products. However, the conception of this approach also implies active engagement of the learners themselves because, "to be able to improve, they must develop the capacity to

monitor the quality of their own work during actual production" (Sadler, 1989, p. 119). Effective formative feedback triggers learners' cognitive mechanisms, pointing to "a gap between actual and desired levels of performance" (Wiliam & Black, 1996, p.543). In addition, formative feedback acts as a cognitive support mechanism since it provides explicit facilitative scaffolding during the learning process. According to Bransford, Brown, and Cocking (2000), scaffolded feedback makes the task more manageable and achievable, as well as indicative of the differences between learners' work and desired outcomes.

- Goal orientation

Feedback on learners' performance can be a powerful motivator, especially if it is provided considering their progress toward a desired goal because the degree of motivation depends on the connection between learners' goals and their expectations regarding achieving those goals (Fisher & Ford, 1998; Ford, Smith, Weissbein, Gully, & Salas, 1998). In psychology, goal orientation refers to individuals' motivation to achieve goals and is distinguished between either performance or learning orientation (Dweck, 1986). Learners with a learning orientation believe that intelligence is ductile and are therefore willing to improve their competence by developing new skills. Learners with a performance orientation, on the other hand, believe that intelligence is innate and thus wish to demonstrate their competence to others and to be positively evaluated by others (Farr, Hofmann, & Ringenbach, 1993).

Fisher and Ford (1998) argue that the learning orientation is more desirable, for it leads to more positive outcomes. It has also been suggested that learners with such goal orientation direct their attention to the task, learn for the sake of learning, and thus devote

greater effort to learning (Dweck, 1986; Button, Mathieu, & Zajac, 1996) by employing complex learning strategies, pursuing challenging tasks, and not giving up in case of failure (Dweck & Leggett, 1988). Formative feedback allows students to verify their expectations and to know how close they are to meeting their learning goals. Therefore, Shute (2008) suggests that formative feedback can influence learners' goal orientations, shifting it from a focus on performance to a focus on learning.

- Complexity of feedback types

The complexity of formative feedback can be realized as verification and elaboration (Kulhavy & Stock, 1989). According to Mason and Bruning (2001) and to Dempsey, Driscoll, and Swindell (1993), verification feedback assesses the correctness of learners' performance in a number of possible ways: *knowledge-of-response*, which tells learners whether their answers are correct or incorrect; *try again*, which informs learners about an incorrect response and allows them more attempts; and *knowledge-of-correct-response*, which provides the correction. Elaboration feedback is more complex in that it elaborates on learners' input by providing relevant hints to direct them towards the expected performance. It can also take different forms. For example, it can be *response-contingent*, focusing on the learner's specific response and describing why something is wrong and why the correction is right; it can consist of *hints/cues/prompts* on how to proceed without providing an explicit correction; it or can be *bug-related* indicating what is wrong and why based on a formal (error) analysis. Most complex is the *informative tutoring* feedback that combines various verifying and elaborating information (Narciss & Huth, 2004). Of all these, the response-specific elaboration feedback was found to enhance student achievement, particularly

learning efficiency, more than other types of feedback (see Mory, 2004). However, research also demonstrates that a combination of verification and elaboration is most efficient (see Mason & Bruning, 2001).

- Specificity

Research has attested that formative feedback leads to correction and is, therefore, more effective when it is specific (Berkenkotter, Huckin, & Ackerman, 1991; Pridemore & Klein, 1995). It should also provide details on how to improve the writing, rather than just indicate deficiencies. Provided with this level of detail, the students are less likely to have negative feedback-related learning experiences or to be confused as to how to respond to the feedback.

- Timing

Formative feedback can be delivered at different points in time during the learning process. *Immediate* feedback is returned instantly after learners finish a given task, and the feedback is *delayed* if it takes a certain amount of time before the learners receive it (which may vary from minutes, to hours, to days). In general, either timing is considered beneficial; immediate feedback is believed to prevent errors from being encoded in the memory, and delayed feedback is thought to reduce the interference factor, allowing the errors to be forgotten. However, instantaneous feedback has been found superior for the acquisition of verbal materials and for procedural skills (Corbett & Anderson, 2001).

Writing can be viewed as a procedural skill in that it entails a mental procedure, which involves the execution of a sequence of mental operations, such as thoughts and decisions regarding the content and the development of a text. Therefore, it may be

speculated that immediate feedback may be more beneficial in developing academic writing skills.

2.3. Computerized intelligent feedback

Practical issues such as large numbers of students, lengthy papers, workload, etc., make it difficult for teachers to provide students with proper feedback (Buchanan, 2000). This is one of the problems in graduate-level academic writing classrooms, where the students come from a variety of disciplines and are required to produce long papers such as research articles, for instance. An even more urgent need in this instructional context is for students to receive iterative, specific feedback pertaining to the writing norms in their disciplines. This need can be met by computerizating formative assessment and feedback. Buchanan (2000) states that learners may even adapt computerized formative assessment to their own learning styles. Not only that; computerization of feedback can also enhance the formative feedback qualities presented above, which are believed to exert positive influences on learning. In knowledge-based courses or in receptive skills language classes, computer-generated feedback of 'pattern markup' and 'error-anticipation' types may be both appropriate and sufficient. For productive language abilities like writing, though, it may be necessary that the feedback is intelligent, or Natural Language Processing (NLP)-generated, in order to reflect the individual-specific problems in learner production.

2.3.1. ICALL systems and feedback

Intelligent feedback has been widely implemented in Intelligent Computer Assisted Language Learning (ICALL) systems. The beginning of 1990s witnessed the piloting of a

number of such systems. For example, *Grammar-Debugger* (Si-Quing & Luomai, 1990) was a parser designed to analyze grammatical as well as ungrammatical sentences produced by Chinese learners of English. Another parser, *Syncheck* (Sanders, 1991; Juozulynas, 1994) analyzed syntax to help learners write German compositions. *GPARS* (Loritz, 1992) included parsers for several languages such as English, Russian, Japanese, and Chinese. *LINGER* (Yazdani, 1991) relied on a combination of tools, e.g., a language-independent parsing mechanism, a language-specific dictionary, and a language-specific grammar, which allowed the system to cope with various aspects of Romance and Germanic languages. Levin, Evans, and Gates (1991) developed a multi-media workbench for learners of foreign languages (*ALICE*), which offered a set of tools for designing different types of intelligent programs in any language. The *ALICE* facilities included tools and resources for syntactic parsing, morphological analysis, error detection, on-line dictionaries and other reference materials, as well as for indexing and retrieval of video/audio/text corpora examples. *ALICE* supported applications targeted at grammar and vocabulary drills, games and simulations, reading and writing assistants, and open-ended learning environments, its two prototypes being a Spanish writing assistant and a Japanese grammar exercise program.

Some early implementations addressed the issue of feedback along with describing the process of ICALL design. *Miniprof* (Labrie & Singh, 1991) was an intelligent tutor for elementary topics in French built to provide grammar exercises, in which the system generated tutorial messages based on the performance of three components - parsing, diagnosis, and tutoring. The function of the parser was to assign a grammatical structure both to the question and to the student's response, and in case the response was an incorrect

sentence, it parsed as much of the sentence as possible to identify the errors. Errors that could not be identified by the parser, typically context and exercise-related kinds of errors, were detected by the diagnostic component. Then, modeling the way in which a teacher would respond when actually tutoring a student, the tutoring component asked the student questions about those errors. Consequently, the program engaged the student in a Socratic dialogue, leading her/him to identify and correct each error (see Fig. 2.2).

#1 Est-ce Que tu gagnes beacoup?
*Non, tu gagnes beacoup.
The question uses the pronoun “tu”.
Which pronoun should you use in your answer?
*?
Use “je”. You need to answer in the negative.
Use the negative in the next sentence.
Answer: Non, je ne gagne pas beacoup.

#2 Est-ce que tu parles espagnol?
*Non, je ne parles espagnol.
The subject is “je”. “parl” is correct.
What should the ending be?
*e
Right. Your answer is negative.
What do you need after the verb?
*I don’t know.
Use “pas”. Answer: Non, je ne parle pas espagnol.

Figure 2.2. An example of *Miniprof*-student dialog (Labrie & Singh, 1991, p. 10)

A parser-based program aimed at checking the grammar in EFL students' compositions was described by Liou (1991). To build the grammar-checker, errors from a collection of student writing samples were found and classified into an error taxonomy, a small electronic dictionary with word stems and necessary features was compiled, a suffix processor to accommodate morpho-syntactic variants of each word stem was designed, after which a parser equipped with phrase structure rules and error patterns was constructed. Once

a grammatical error was detected, the program output a message to the learner by employing a message generating routine, in which a flag attached to each processing rule was matched with a message file. The feedback message was created from a template and consisted of the erroneous sentence, an explanation, and possibly examples (see Fig. 2.3).

Error:
The *development* in scientific technologies *have* bring great change.
Feedback:
[*development* is the subject of the verb *have*. The subject is in 3rd person singular form.
The following are 2 correct examples:
The *clerk* beside the bookshelves *is* watching television.
The *lady* who the workers *love* teache*s* English.]

Figure 2.3. An example of the grammar-checker program feedback (Liou, 1991, p. 66)

Another program that revolved around a parser was *BRIDGE* (Holland, Maisano, Alderks, & Martin, 1993), a German tutor which analyzed learners' responses in the form of free sentences. It was developed for the use of military linguists in the US Army. The interface included a given text, graphical aids such as maps, learner aids such as a dictionary, and lesson exercises which consisted of oral or written questions to which the learner responded with a written sentence. Upon the reception of the response, the program passed the sentence to the parser, and the tutor informed the learner of a primary error and offered a chance for correction. For example, in response to a student's syntactically errorful '*Lauterbach liegen nordoestlich von das Tal*', the system identified as a primary error the case of the noun phrase, expressed in 'das', accompanying the preposition 'von' and generated a feedback message as shown in Figure 2.4.

PRIMARY ERRORS [1/1]
Prep noun error: von, das

Figure 2.4. A sample of *BRIDGE* feedback (Holland et al., 1993, p. 33)

Secondary errors were not automatically displayed. The learner was alerted of a secondary error, but had the choice of whether to look it up or not. If the revised response was grammatical, the tutor returned a message to that effect. Holland et al. (1993) claim that their systems' "feedback can be couched in pragmatically relevant terms. For example, besides saying 'verb and subject don't agree', as in *BRIDGE*, we could add 'This might confuse the reader/hearer as to whether you mean singular or plural' or 'This kind of mistake will lead people to doubt your competence in the language'. Thus, feedback could call attention to the public consequences of grammatical acts." (p. 41)

NIHONGO-CALI, a Japanese Computer Assisted Language Instruction program, (Nagata, 1993) provided exercises for Japanese passive structures, which learners could complete by inputting full sentences. It employed NLP to facilitate error analysis of responses in the light of morphological and syntactical rules of the Japanese language and to generate feedback. At the core of the program was an NLP analyzer, which consisted of a morphological parser, a syntactic parser, and a knowledge base with three components: a core lexicon, a set of morphological rules, and a set of grammar rules. Capable of parsing both grammatical and ungrammatical input, when a submitted sentence was incorrect, *NIHONGO-CALI* generated an output error message, which was in turn sent to the pattern matching program to verify whether the learner's response was appropriate for the question in the given exercise. Depending on the pattern matching result, the learner received feedback in terms of grammatical functions and thematic roles of each sentence constituent (see Fig. 2.5). Several errors could be detected by the NLP analyzer and the pattern matching program, the errors being classified into vocabulary, particle, verbal predicate, and nominal modifier

errors. The system was also teacher-friendly in that it allowed the instructors to create feedback messages themselves, depending on their instructional focus.

Options

1. Respond to the following conversation according to the situation described.

(1) Your colleague told you that he called your office this morning, but nobody answered. Respond that yes, the secretary came in late (i.e., you were affected by the secretary's coming in late).

Colleague: Kesa ztnusyo nidenwa stnastta kedo, dare mo demasen destta yo
Your response: Htsyo ga osoku ni koraeta.

Read the following messages:
<Vocabulary error>
- EE or HAI is missing (trivial error)
<Particle error>
- GA is not expected to be used here.
- In you sentence, HISYO is the 'subject' of the passive (the one that is affected by the action), but it should be the 'agent' of the passive (the one who performs the action and affects the subject). Use the particle NI to mark it.
- OSOKU modifies the verbal predicate without the particle NI.
<Verbal predicate error>
- Since you are talking with a colleague, use the distal-style (-masu form).

Figure 2.5. An example of *NIHONGO-CALI* feedback (Nagata, 1993, p. 333)

Yang and Akahori (1998) developed a web-based system which, similar to *NIHONGO-CALI*, accepts free response sentences in Japanese, the difference being that the former worked with the Latin alphabet, while the latter can analyze Japanese characters. Major components of the system include: an interface (i.e., a WWW browser); a sentence analysis system with a morpheme analyzer, a syntax analyzer, and an error analyzer; grammar rules; a feedback system with a message generator, a knowledge database, and a list of students' learning histories; and a dictionary. The learner's typed in sentences are processed by the morpheme analyzer, and the sentence is checked by a case grammar. If an

error is found, the grammar rules are applied to identify the type of error and the order of priority so that a relevant feedback message is created. The sentence analysis component sends an error ID to the feedback component, which determines the kind of feedback message to be displayed to the learner. This message indicates the type of error and points to its location rather than correcting the error, therefore directing the learner to attempt self-correction. The learner can choose to correct her/his sentence immediately, or to access grammatical explanations first and correct the sentence later (see Fig. 2.6). An advantage of this program is the fact that the feedback is displayed not only depending on the type of error/s identified, but also depending on the learner who submits the sentence. This adaptive individualized approach to feedback generation, although still very limited, is made possible by the web-based interface, which requires learners to enter their name and information about their experience with the Japanese language when logging in.

Input 1: *Kaku san wa Jyon san o nagurimashita*. 'Kaku hit John.'

Type of error ID: ERR_A11

Feedback Messages:

Hello, [student's name]

Input Japanese sentence is as follows:

"*Kaku san wa Jyon san o nagurimashita*."

In this sentence, part of "*nagurimashita*" has a mistake in 'the verb form. The passive form should be used instead of the active form.'

Click "Examples" to see examples of passive form.

Please modify your answer or refer to the links below.

Figure 2.6. Examples of the web-based Japanese program feedback (Yang & Akahori, 1998, pp. 60-61)

Links:

"Table of verbs"

What is a passive form?

Examples of passive forms

How to make a passive form

Input 2: *Jyon san wa Kaku san ni naguriraremashita*.

Type of error ID: ERR_A31

Feedback Messages:

Hello, [student's name]

Input Japanese sentence is as follows:

"*Jyon san wa Kaku san ni naguriraremashita*."

> In this sentence, part of "*naguriraremashita*" has a mistake in 'the conjugation of the verb. You should convert the verb to write it in the passive form.'

Click "Examples" to see examples of the passive form.

Please modify your answer or refer to the links below.

Links:

"Table of verbs"

What is the passive form?

Examples of passive forms

How to make a passive form

Input 3: *Jyon san wa Kaku san ni naguraremashita*. 'John was hit by Kaku.'

Type of error ID: ERR_X9

Feedback message:

Congratulations!!! Your answer is right.

Go on to another question.

Figure 2.6. Examples of the web-based Japanese program feedback (Yang & Akahori, 1998, pp. 60-61) (continued)

Another web-based system, the *German Tutor* (Heift, 2002, 2003; Heift & Nicholson, 2001), is an intelligent language tutoring system which, similar to other programs described above, analyzes sentence-level input and detects language errors. It contains a domain knowledge that represents the knowledge of language encoded in the grammar, an analysis module that yields possible responses to the learner's input, a filtering module that makes decisions regarding the order of feedback, and, superior to the functionality of other programs, it has an adaptive Student Model, which maintains a record of individual learners' performance. Heift explains that "performance over time is monitored across different grammatical constructs; the information is used to tailor feedback messages suited to learner expertise within a framework of guided discovery learning" (p. 298). The grammar parser's detailed output is used to select error-specific feedback messages, which differ depending on the level of language competence. Beginners receive explicit messages while advanced learners are provided with a hint at the error expressed in more metalinguistic terms (see Figures 2.7a, 2.7b, 2.7c, and 2.7d). In the case of multiple errors, the *German Tutor* follows the principle of error prioritization, thus its feedback treats each error one at a time. Since this is a tutoring system, it provides recommendations for remediation in addition to the feedback by suggesting additional exercises depending on the number and type of errors that have been detected.

Feedback message for a beginner student:

(1a) *Bianca hat ohne er gegangen.
(1b) Bianca ist ohne ihn gegangen.
(1c) Hier stimmt das Hilfsverb HAT nicht.
(HAT is wrong here.)
(*GEHEN* requires *IST.*)

Figure 2.7a. An example of *German Tutor* feedback (Heift, 2002, p. 298)

Feedback message for an intermediate student:

(1a) *Bianca hat ohne er gegangen.
(1b) Bianca ist ohne ihn gegangen.
(1c) Hier stimmt das Hilfsverb HAT nicht.
(The auxiliary is wrong here.)

Figure 2.7a. An example of *German Tutor* feedback (Heift, 2002, p. 298) (continued)

Similar intelligent language tutoring systems with the same feedback features have been developed for English (Toole & Heift, 2002) and Greek (Heift, Toole, McFetridge, Popwich, & Tsiplakou, 2000).

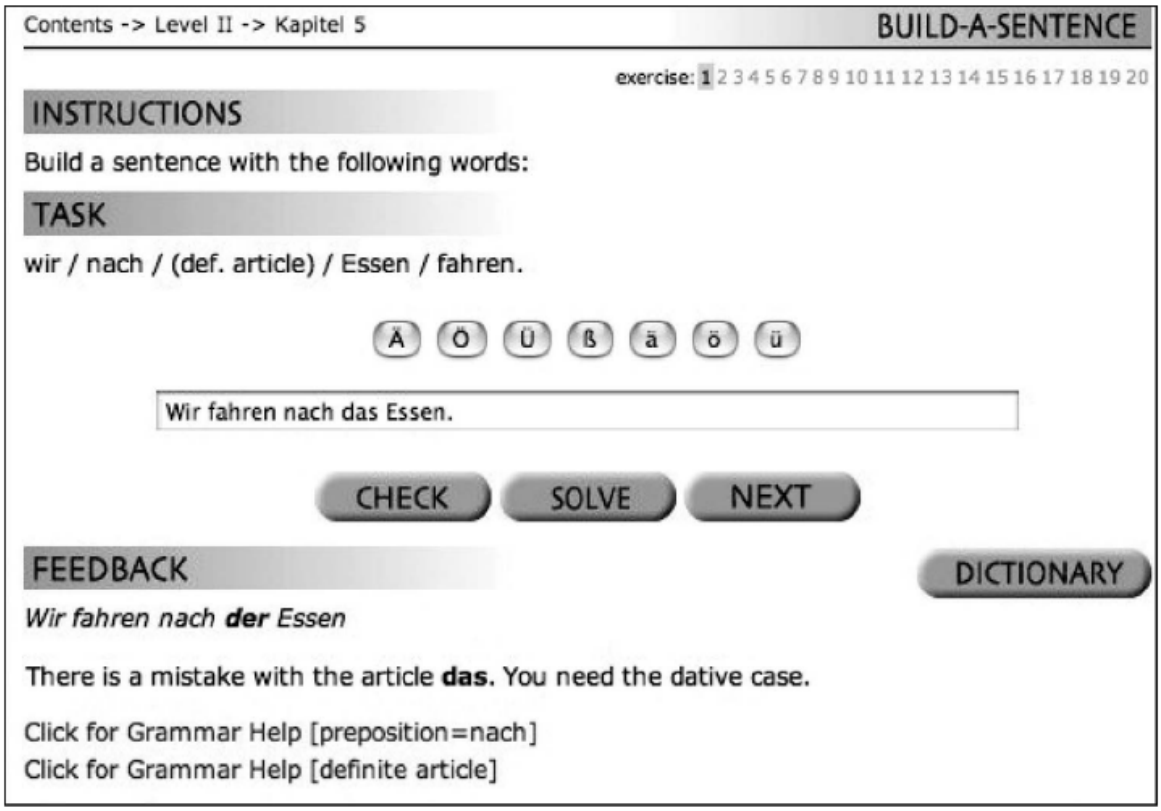


Figure 2.7b. An example of *E- Tutor* feedback (Heift, 2006, p. 246)

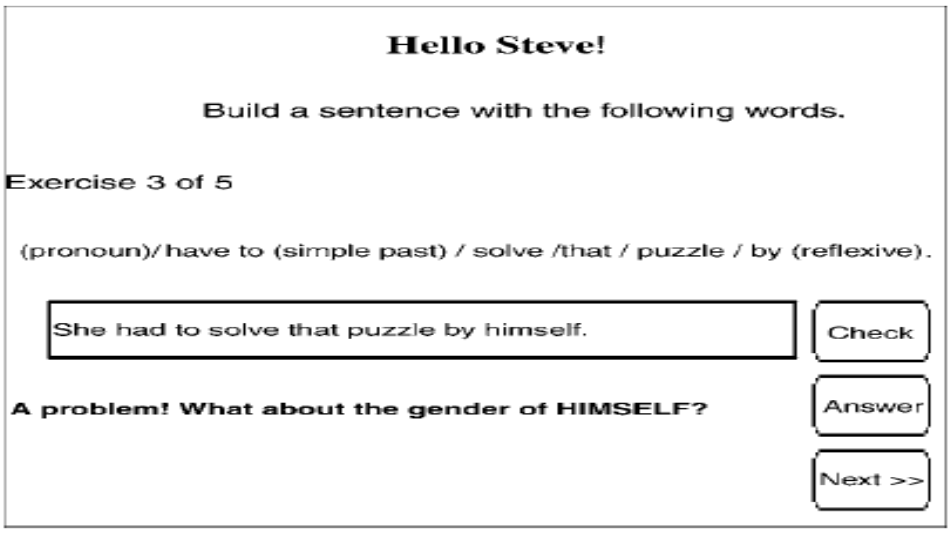


Figure 2.7c. An example of *ESL Tutor* feedback for a beginning language learner (Toole & Heift, 2002, p. 376)

The Student Model in these systems keeps track of various error types such as grammar, vocabulary, punctuation, etc., and at the end of every exercise set, learners receive a detailed summary of the errors they have made. The summary page in Fig. 2.7d states that the learner made one spelling mistake and ten errors in subject-verb agreement and advises him to further complete exercises on subject-verb agreement.

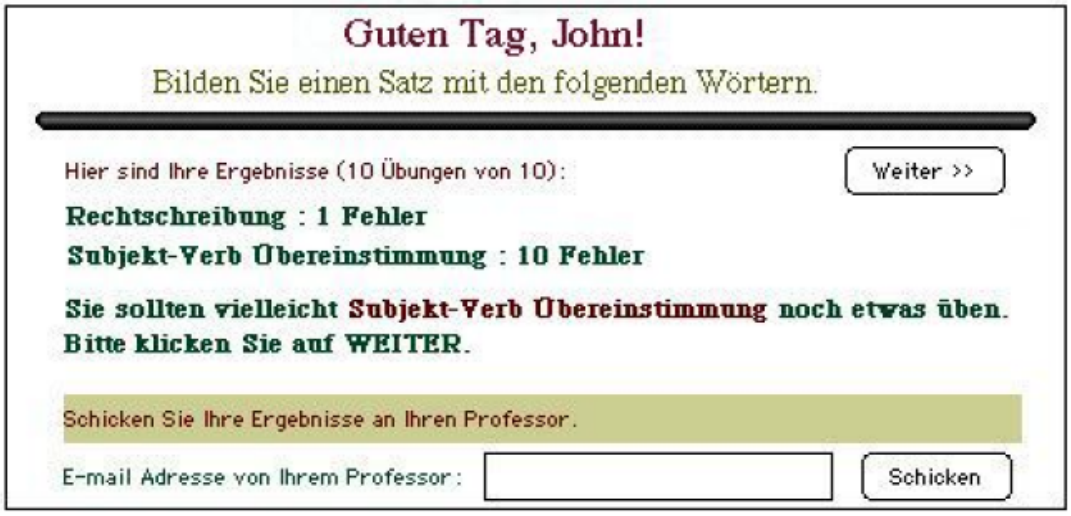


Figure 2.7d. An example of *German Tutor* error summary (Heift & Nicholson, 2001, p. 318)

Azalea is an interactive intelligent language tutoring system developed for an ESL composition course (Chen & Tokuda, 2003; Chen, Tokuda, & Xiao, 2002; Tokuda & Chen, 2001, 2004). It consists of a template automaton structure for knowledge representation, a diagnostic engine based on a matching algorithm, a part-of-speech tagged parser for syntactic analysis, a parser-based learner model, and a visual interface authoring tool. The error messages provided through template matching result from a total of 164 errors that the system can diagnose, these errors being classified as grammatical errors, word and usage errors, and non-preclassified errors. Feedback messages are returned depending on whether the learner's response finds an exact match with the model translations provided or with the finite states of the template, on whether the learner's response does not exactly match the finite states of the templates (2.8a, 2.8b), or on whether the learner's response matches less than half of the template path (2.8c).

Model translation: In recent days, many different models have become available.
Input sentence: In the recent days, many different model is become available.
Diagnosis given: In [the](1) recent days, many different [model](2) [is](3) become available.
Comment (1): When referring to general classes of things, do not use "the" before plural countable nouns (boats, motors, meetings, countries). This part is not needed.
Comment (2): This noun should be plural. This part should be: models
Comment(3): This part should be: have

Figure 2.8a. An example of *Azalea* feedback (Tokuda & Chen, 2004, p. 196)

Model translation 1: The work on the new material hasn't been a success.
Model translation 2: The research into the new materials ended in failure.
Model translation 3: Our research into new materials went unsuccessfully.
Model translation 4: We were unsuccessful in our studies on the new materials.
Input sentence: The research onto new materials failure in ended.
Diagnosis given: The research [onto](1) new materials [XXX](2) failure [in ended](3).
Comment (1): To translate the phrase ～に関する研究, use "the research into ～" or "the research on ～." This part should be written as follows: into / about / concerning / on / regarding
Comment (2): Something is missing. This part should be: brought
Comment (3): This part is not needed.

Figure 2.8b. An example of *Azalea* feedback (Tokuda & Chen, 2004, p. 197)

Model translation 1: Galileo's theory exerted a major influence on physics.
Model translation 2: Physics was greatly affected by Galileo's theory.
Model translation 3: Physics was influenced by Galileo's theories to a great extent.
Input sentence: Galileo's theory has big Physiks influence.
Diagnosis given: Too Many Errors, Try again.

Figure 2.8c. An example of *Azalea* feedback (Tokuda & Chen, 2004, p. 197)

Reuer (2003) reports on an ICALL system that invites learners to use language interactively in order to enhance the development of communicative competence. It provides small question-answer dialogs with the computer during which learners are required to produce complete written sentences. The parser described in this article relies on Lexical Functional Grammar (LFG) and uses an unrestricted grammar and lexicon. It contains some linguistic knowledge from an analysis of a learner corpus of German to constrain the search space, but it does not include any buggy or mal-rules for describing erroneous sentences (see Figure 2.9).

I has seen an accident.

I has seen an accident.

$$\begin{bmatrix} \text{auxtype}: \text{have} \\ \text{fin}: + \\ \text{pred}: \text{SEE(subj, obj\_dir)} \\ \text{subj}: \begin{bmatrix} \text{case}: \text{nom} \\ \text{num}: \text{sg} \\ \text{pers}: \text{I} \\ \hline \textit{err}: [\text{pers}: \text{III}] \end{bmatrix} \\ \text{obj\_dir}: \begin{bmatrix} \text{case}: \text{acc} \\ \text{num}: \text{sg} \\ \text{pred}: \text{ACCIDENT} \end{bmatrix} \\ \hline \textit{err}: [\text{subj}] \end{bmatrix}$$

A possible feedback message to learners might be: "There is an agreement error between the subject and the verb. The subject should be third person, but you chose first person." As mentioned above, the phrase structure tree can be used for the generation of a correction of the sentence. Moving downward in the tree tells us where the error originated, and the correct forms can be retrieved from the lexicon.

Figure 2.9. An example of the interactive ICALL system feedback (Reuer, 2003, pp. 508-509)

Although Reuer (2003) does not directly tackle the issue of individualized feedback, he emphasizes the possibility of giving LFG-type feedback if a syntactically ill formed sentence

is encountered, especially in traditional or pedagogical grammar teaching where the terminology would be adequate and comprehensible to students. The author argues that parser-generated structures are suitable for constructing error messages with little need for additional explanation.

Arabic ICALL for learning the Arabic language was designed by Shaalan (2005). It has two main types of test items for interaction with the learner: selection-type (matching, true/false, identify, or multiple-choice) interactive questions supply-type or (short-answer/fill-in-the-blank). Like other intelligent systems, it encourages free production of sentence-length input and guides learners towards the recognition of erroneous, inappropriate, or misused expressions. The overall architecture of this application includes such modules as: user interface, course material (with educational units, an item bank, a test generator, and an acquisition tool), sentence analysis (with a morphological analyzer, a syntax parser, grammatical and 'buggy' rules, and a lexicon), and the feedback module. The last module contains an error analyzer that parses ill-formed input, compares the analysis of the learner's answer with the correct answer generated by the system, and issues feedback. (Fig. 2.10 shows a sample analysis of a possible learner's answer and the possible corresponding feedback.) The learner can either correct the error immediately, or s/he can restudy the related grammar items first and then attempt to make a correction.

Task:

Complete the sentence with the correct unrestricted object.
المفعول المطلق"
أكمل بمفعول مطلق مناسب:
. •أبر أبى _____. (. _____.I am kind to my father)

Figure 2.10. An example of *Arabic ICALL* feedback (Reuer, 2003, pp. 508-509)

Possible learner's answer along with the corresponding feedback:

- A word that is not a noun. Issue a message describing that the unrestricted object should be a noun.
- A word that is a noun but does not originate from the infinitive verb. Issue a message describing that the unrestricted object should be the infinitive of the verb.
- A word that is both a noun and originates from the infinitive verb but is defined. Issue a message describing that the unrestricted object should be undefined.
- A word that is a noun, originates from the infinitive verb but needs the end case ''Alef Tanween'', and is undefined. Issue a message describing that a missing end case of the unrestricted object.
- A Correct answer. Issue a positive message.

Figure 2.10. An example of *Arabic ICALL* feedback (Reuer, 2003, pp. 508-509) (continued)

Existing ICALL programs have a lot in common. They all seem to target the development of learners' grammatical competence, revolve around the functionality of a parser, and provide immediate intelligent feedback. Their feedback features display great instructional potential, varying in their degree of specificity and explicitness as well as in their ability to adapt to individual learners (see Table 2.1). The question, however, is not what kind of feedback ICALL can generate; rather, it is what kind of feedback ICALL programs should generate. In other words, what intelligent feedback should be given in order to facilitate second language acquisition? Nagata (1995) argues that "[i]f we use an intelligent system, we should examine carefully what kind of error messages should be provided […] and how effective they are" (p. 49).

Table 2.1. Overview of ICALL systems

ICALL system	*System components*	*Language*	*Skills*	*Type of feedback*
Labrie & Singh (1991) - *Miniprof*	parser, error diagnosis, tutor	French	grammar	dialog, error-specific, explicit pointer to error, metalinguistic
Liou (1991)	pattern matching, parser, message generator	English	grammar	error-specific, explicit pointer to error, metalinguistic
Holland et al. (1993) - *BRIDGE*	parser, graphical aids morphological parser, syntactic parser, core lexicon, morphological	German	grammar (spatial description)	examples right/wrong, error-specific, explicit pointer to error
Nagata (1993) - *NIHONGO-CALI*	rules, grammar rules	Japanese	Grammar (passive structures); vocabulary	error-specific, explicit pointer to error, metalinguistic
Chen & Tokuda (2003), Tokuda & Chen (2001, 2004) – *Azalea*	template automaton structure, diagnostic engine, POS parser, parser-based learner model, visual interface	English	grammar, vocabulary	error-specific, explicit pointer to error, metalinguistic, correction
Reuer (2003)	LFG parser, unrestricted grammar and lexicon, linguistic knowledge	English	grammar	error-specific, explicit pointer to error, metalinguistic, tree-form
Shaalan (2005) - *Arabic ICALL*	user interface, course material, morphological analyzer, syntax parser, non/grammatical rules, lexicon, feedback module	Arabic	grammar	error-specific, explicit pointer to error, metalinguistic
Yang and Akahori (1998)	WWW interface, morpheme analyzer, syntax analyzer, analyzer; grammar rules, feedback system, knowledge database, students' learning histories, dictionary	Japanese	grammar	error-specific, explicit pointer to error, metalinguistic, examples, source links, adaptive individualized
Heift (2002, 2003)	domain knowledge, analysis module, filtering module, Student Model	German	grammar	error-specific, level-tailored, explicit pointer to error, metalinguistic, error-prioritized, remediation exercises, adaptive individualized, error summary

As it can be seen in Table 2.1, ICALL feedback has been widely employed to assist students with grammar. According to previous research, such is effective if it is: individualized according to specific learner input, pointing to the error type, explicitly explaining the error, iterative, and leading to self-correction.

When comparing intelligent feedback, (detailed grammatical explanations of the learners' errors) with CALL feedback (messages only pointing to missing or unexpected words in the learners' responses), Nagata (1993) found that the former is more effective than the latter. Moreover, intelligent feedback appears to be more effective even than the enhanced version of traditional feedback, which also indicates the positions of the missing items (Nagata, 1995). Further, Yang and Akahori's (1999) findings indicate that feedback which corresponds to the input created by the learner is superior to feedback displaying the correct answer in a 'multiple selection' method in that it enhances self-correction. In terms of correction, van der Linden (1993), while examining the strategies learners employed when interacting with different levels of feedback, observed that learners felt motivated to self-correct when they received feedback about the type of error committed.

The more detailed the feedback is, the better the learning outcomes are (Heift, 2001, 2002, 2004, 2005); therefore, the feedback should not only display the error, but also explain why the response is inappropriate. Heift (2008) explored the effects of two types of feedback, metalinguistic explanation and metalinguistic clue, on beginner, advanced beginner, and low intermediate learners of German, looking into the longitudinal effects of these types of feedback and into their impact on grammar and spelling error classes. Overall, she concluded that learners show significantly more learner uptake over time with the more error-

specific/explicit feedback type. In another study (Heift & Rimrott, 2008), the researchers considered three distinct feedback types for spelling errors: meta-linguistic with emphasis (feedback includes suggestions for the misspelled word in addition to displaying the incorrect sentence and highlighting the error), meta-linguistic (feedback indicates that a spelling error has occurred and provides a list of suggestions), and repetition (feedback does not contain any suggestions for correction). Here, too, learners had the most correct responses and the greatest uptake with the most explicit and prominent feedback type, i.e., meta-linguistic with emphasis, repetition being the least effective.

2.3.2. Conceptualization and operationalization of computerized formative feedback

As demonstrated above, encouraging efforts have been made in the area of computerized feedback on grammar. While informative, they are not entirely elucidative for computer assisted formative assessment in second language writing, for which no research has investigated the effectiveness of ICALL feedback. What is also notably needed is conceptualizing and operationalizing the process of feedback through a more complex framework that would place a significant emphasis on students' production.

A suitable framework for conceptualizing and operationalizing computerized formative feedback is Evidence Centered Design, which provides a conceptual framework "for the elements of a coherent assessment, at a level of generality that supports a broad range of assessment types, from familiar standardized tests and classroom quizzes, to coached practice systems and simulation-based assessments, to portfolios and student-tutor interaction" (Mislevy, Almond, & Lukas, 2003, pp. 3-4). This framework relies on

evidentiary arguments and, according to Almond, Steinberg, and Mislevy (2002), it "describes a process that begins by defining the decisions to be made based upon the results of an assessment and then works backwards to develop tasks, delivery mechanisms, scoring procedures, and feedback mechanisms that provide evidence that informs the pre-defined purposes" (p.4). ECD relies on Conceptual Assessment Framework (CAF) models for design and on a generic framework for assessment delivery (Mislevy, et al., 2003). The CAF contains six models, which serve as a blueprint for the operational elements of the assessment and function as follows. The 'Student Model' defines the knowledge, skills, and abilities to be measured. The Evidence Model supplies instructions on how to interpret the performance in the form of examinees' work products, i.e., their production responses to given tasks. The Task Model delineates the tasks that can elicit the kinds of evidence needed for the evidence model. The Assembly Model controls the selection of tasks for the creation of an assessment. The Presentation Model specifies how a particular task is to be presented in a given delivery environment. The Delivery System Model describes how student, evidence, task, assembly, and presentation models work together.

Since delivery of assessments is possible in different ways, e.g., paper and pencil, oral, computer-based, etc., ECD provides a generic Four Process Delivery Architecture. These four processes, shown in Figure 2.11, are all necessary to make observations and draw inferences that form an assessment argument. First, a task is selected by the Activity Selection Process, then it is sent for display to the Presentation Process. Once the student finishes completing the task, his/her work product is forwarded to the Response Processing, which identifies relevant evidence in the work product and transfers it to the Summary

Scoring Process. Finally, the Activity Selection Process makes a decision about the next step to be taken.

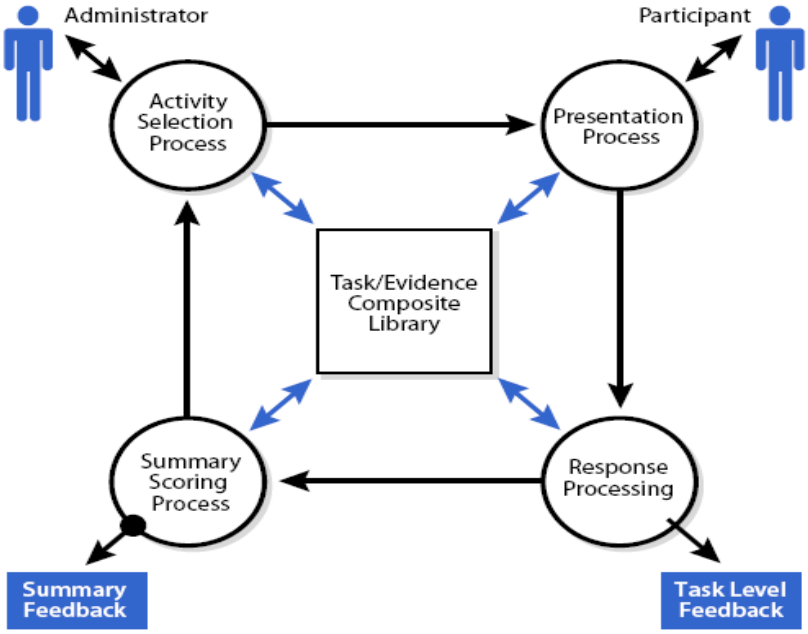


Figure 2.11. The four principle processes in the assessment cycle (from Almond et al., 2002, p. 7)

In sum, ECD, with its underscoring of evidence obtained from learners' work products, constitutes a valid conceptualization for the process of providing computer-based formative feedback and can be applied to the teaching of academic writing. Similarly, the ECD delivery mechanisms can inform the operationalization of the feedback process. What appears to be particularly promising for computer-based writing instruction is ICALL formative feedback generated by automated writing evaluation (AWE) systems, which are capable both of performing analyses of learners' production and of incorporating the assessment delivery elements, thus satisfying both conceptualization and operationalization specifications for formative feedback.

2.4. Automated Writing Evaluation (AWE)

2.4.1. AWE systems

Automated writing evaluation (AWE), also referred to as computerized essay scoring, computer essay grading, computer-assisted writing assessment, or machine scoring of essays is defined as "the ability of computer technology to evaluate and score written prose" (Shermis & Burstein, 2003, p. xiii). Investigations of AWE originated in the mid 1960s (Page, 2003). Since then, due to significant advances in computing technology, the potential of AWE has been largely deployed in research, which has been informed by multiple perspectives from teaching pedagogy, educational measurement, cognitive science, and computational linguistics. In other words, what is considered to be most beneficial for students, models which reflect students' thought processes, psychometric evaluations of reliability and validity, considerations about operational systems and their functionality - all have contributed to the development and implementation of AWE systems.

These systems have now become an important supplement for summative, large-scale assessment contexts. Performance-based and high-stakes standardized tests assess constructed responses, which are automatically evaluated in terms of topical content, grammar, style, mechanics, syntactic complexity, etc. For example, ACT, GMAT, and TOEFL use the E-Rater for this purpose, as shown in Table 2, which also lists other scoring systems and targeted constructs.

Table 2.2. Automated essay scoring systems and assessed constructs

Automated essay scoring system	*Constructs*
Project Essay Grade (Page, 2003)	Grammar, fluency, diction
Intelligent Essay Assessor (Landauer et al., 2003)	Content; Grammar, style, mechanics Plagiarism and deviance

Table 2.2. Automated essay scoring systems and assessed constructs (continued)

Automated essay scoring system	*Constructs*
E-Rater (Burstein, 2003)	Plagiarism and deviance Topical content Rhetorical structure Syntactic complexity
BETSY (Rudner and Liang, 2002)	Content Grammar, style, mechanics
IntelliMetric (Elliot, 2003)	Focus / unity Development / elaboration Organization / structure Sentence structure Mechanics / conventions
Automark (Mitchell et al., 2002)	Content Grammar, style, mechanics

However, while various NLP and statistical techniques at the basis of automated scoring engines allow for the evaluation of a number of writing constructs, when it comes to text organization, this evaluation is limited to recognizing the five-paragraph essay format with its thesis and topic sentences. As of now, there are no AWE systems that extrapolate the technological possibilities to other academic genres, such as the research reports for instance, which are a major challenge for new NNS members of academia.

Nevertheless, because AWE systems are able to capture all these peculiarities of student writing, they are beginning to draw the attention of pedagogues who are very interested in assessment for learning, i.e., assessment used as a tool in gaining direction for remediation. Programs such as *Criterion*^{SM} by Educational Testing Service, WriteToLearn by Pearson, and *MyAccess!* by Vantage Learning, have widely been used in writing classrooms with the purpose of enhancing the learning opportunity by allowing students to submit their essays, receive feedback, improve and re-submit their texts. Table 3 briefly summarizes the most essential information about their origin, abilities, and instructional contexts.

Table 2.3. AWE programs

	Criterion	*MY Access!*	*WriteToLearn*
Developer	Educational Testing Service	Vantage Learning	Pearson Education
Scoring engine	e-rater	Intellimetric	Intelligent Essay Assessor
Purpose	Development of writing skills	Development of writing skills	Development of reading comprehension and writing skills
Analysis and feedback	Syntax, discourse, topical content, lexical complexity, grammar, usage, mechanics, style	Focus and meaning, organization, content and development, language use and style, mechanics	Ideas, organization, conventions, sentence fluency, word choice, writer's voice, spelling, copying, redundancy, irrelevancy
Context	K-12, colleges, US and foreign universities, national job training programs, military institutions	K-12, colleges, US and foreign universities	K-12, colleges, US universities

All these AWE programs include a wide variety of complementary instructional tools that are intended to assist both students and teachers. For students, these programs display automated as well as teacher feedback, which can be solicited and received at different stages of writing. This helps students to focus not only on automatically detectable errors, but also on other, more subtle, aspects of writing. They can also view their performance summary that typically includes a holistic score, the number of errors, and links to detailed feedback on each error category. In addition, *Criterion* has a context-sensitive Writer's Handbook (also available in a Spanish-English bilingual version) that provides additional definitions and lessons. To assist students in their planning process, this program offers a 'Make a plan' tool

with a choice of eight templates for planning strategies. *MY Access!*, in turn, has an online writing coach, which evaluates student writing and provides revision goals and remediation activities for each of the writing traits, as well as an editor, which highlights errors and provides editing suggestions. It also offers a writer's checklist for guidance, scoring rubrics for self-assessment, word banks for appropriate vocabulary use, and graphical pre-writing tools for better formulation and organization of ideas. *WriteToLearn* has similar options; plus, it allows students to hear the text in reading passages (through text-to-speech technologies) and to see the translation of words and their dictionary definitions (through pop-up windows) on demand.

These three programs allow teachers to create their own writing assignments or to select topics from the system's library. In *MY Access!*, teachers can choose from a large pool of over 700 prompts, including narrative, persuasive, informative, literary, and expository genres. In order to make it an integrated writing instruction tool, the program's prompts were selected in line with popular textbook series and state standards. *WriteToLearn* has a comprehensive database of more than 300 reading passages in science, history, language arts, social studies, and fiction, targeted at students in grades 4 through 12. Moreover, teachers can set certain parameters for their assignments. Additionally, all these programs have powerful data-analysis and reporting features that allow for organizing student essays, storing them for quick retrieval, and generating error reports. The reports can also be at the level of performance summary, group comparison, student history, etc., allowing teachers to monitor student progress individually and as a class. Structured online training modules are available to assure an improved instructor experience.

2.4.2. AWE for teaching and learning

This brief overview reveals a number of advantages of AWE for classroom use. It should be mentioned that all the AWE advantages considered here are technically possible, but not yet empirically supported. Hyland and Hyland (2006) point out that research has not been unequivocally positive about the role of formative feedback in L2 writing development since many pedagogical issues regarding feedback remain only partially addressed. That is why it is important that AWE be implemented in a principled way, taking into account teaching and learning needs and goals as well as theoretical insights and research findings. In addition, the actual impact of automated feedback on the development of writing skills has not yet been clearly understood (Warschauer & Ware, 2006). Therefore, the claims made in this section of the chapter with regards to the potential of automated evaluation to enhance the formative feedback qualities are hypothetical at this point. Future investigations of the AWE effects on learning, such as the one in this dissertation work, may uncover evidence that could confirm these tentative claims.

The technical capabilities of the automated scorers of these programs allow for generating feedback on a wealth of writing traits, which, as AWE supporters argue, can dramatically ease the burden on teachers. *Criterion*'s feedback, though, seems to be the most complex. It is also better because, unlike *MY Access!*, *Criterion* can provide feedback even in response to prompts that have not previously been rated by human scores. *WriteToLearn* provides feedback that is based on an inference of semantic similarity of words and passages based on analyses of large bodies of relevant text.

Second, all the features incorporated in these AWE programs have been claimed as effective supplements to instruction by their developing companies, which maintain that they can potentially guide students through their writing process, help them self-assess their work and monitor their progress, and thus possibly enhance learner autonomy. Developers also assert that, because these programs not only allow for, but also facilitate student-teacher interaction, the value of human audience is not meant to be eliminated, and teacher's evaluative influence is expected to remain strong. As for teachers, the wide selection of topics aims at helping the teachers find some content of the writing prompts that would be fit with their curricular needs. Finally, teachers' access to their students' individual portfolios is alleged to be useful not only to evaluate progress, but also to reinforce discussions with students, administrators, and parents. These are positive, but only judgmental suppositions put forth by the developers of AWE programs, which may be impressive and sufficient for marketing, but not entirely convincing for implementation, which requires empirical evidence about the impact of AWE on learning and instruction.

Another advantage of AWE is that this technology opens wide possibilities for high quality formative assessment by helping match teaching goals and tailor writing instruction to individual learners, such as those in the academic writing courses targeted by this study. AWE feedback can be considered formative particularly because it can possess the qualities described in Section 2.2.2: cognitive involvement, goal-orientation, complexity, specificity, and timing. Judgmental considerations of how automated feedback can meet the specifications of these qualities are discussed below.

Feedback generated by a natural language processing component of an AWE program can objectively and accurately indicate the problems contained in students' writing (given that the automated analysis is reliable and valid), helping them make corrections, or in other words, resolve the 'gaps'. In addition, it may help reduce students' uncertainty about how good (or poor) their writing is, which is often caused by the delayed feedback provided by the teacher. Consequently, automated feedback can potentially stimulate higher levels of cognitive involvement and writing effort. Furthermore, AWE systems can be programmed to provide scaffolding (Vygotsky, 1987) in the form of feedback that would include cues, prompts, tips, or partial solutions for individual students, depending on the purposes and needs of writing instruction.

An AWE program's feedback can be considered goal-oriented because, by allowing students to view the evaluation of and feedback on their performance at any stage in the writing process, it offers them opportunities to self-assess their progress toward the goal they want to achieve while completing the given writing task. The intelligence of the automated evaluation system can generate feedback that can clearly indicate the students' current performance characteristics compared to the desired output. An example of such feedback is provided in Figure 2.12. Overall, by returning goal-orientation feedback, AWE may lead to plausible effects such as, for example, an increase in students' understanding that the writing skills can be developed through practice, that failures are part of the skill-acquisition process, and that effort is critical in developing these skills (Hoska, 1993).

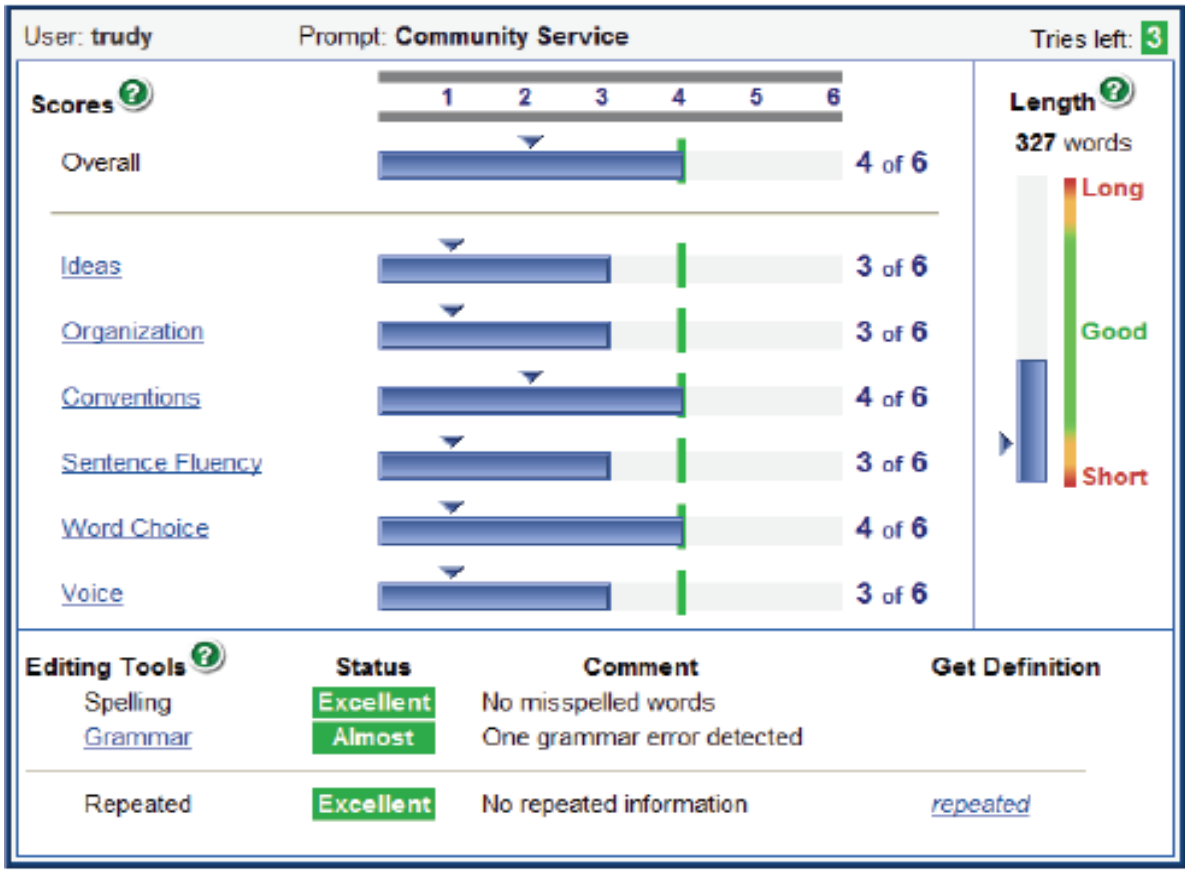


Figure 2.12. Example of goal-oriented feedback in WriteToLearn

In addition, AWE can generate complex feedback, including both verification (Figure 2.13) and elaboration (Figure 2.14) feedback types, which is not always feasible or possible for CALL programs or instructors. It can be facilitative and directive at the same time, evaluating the correctness of learners' responses, highlighting particular errors, indicating their source, recommending corrections, and offering guiding suggestions. The fact that AWE feedback is instant is a feature that may increase its formative helpfulness when provided for the purpose of developing a procedural skill like writing.

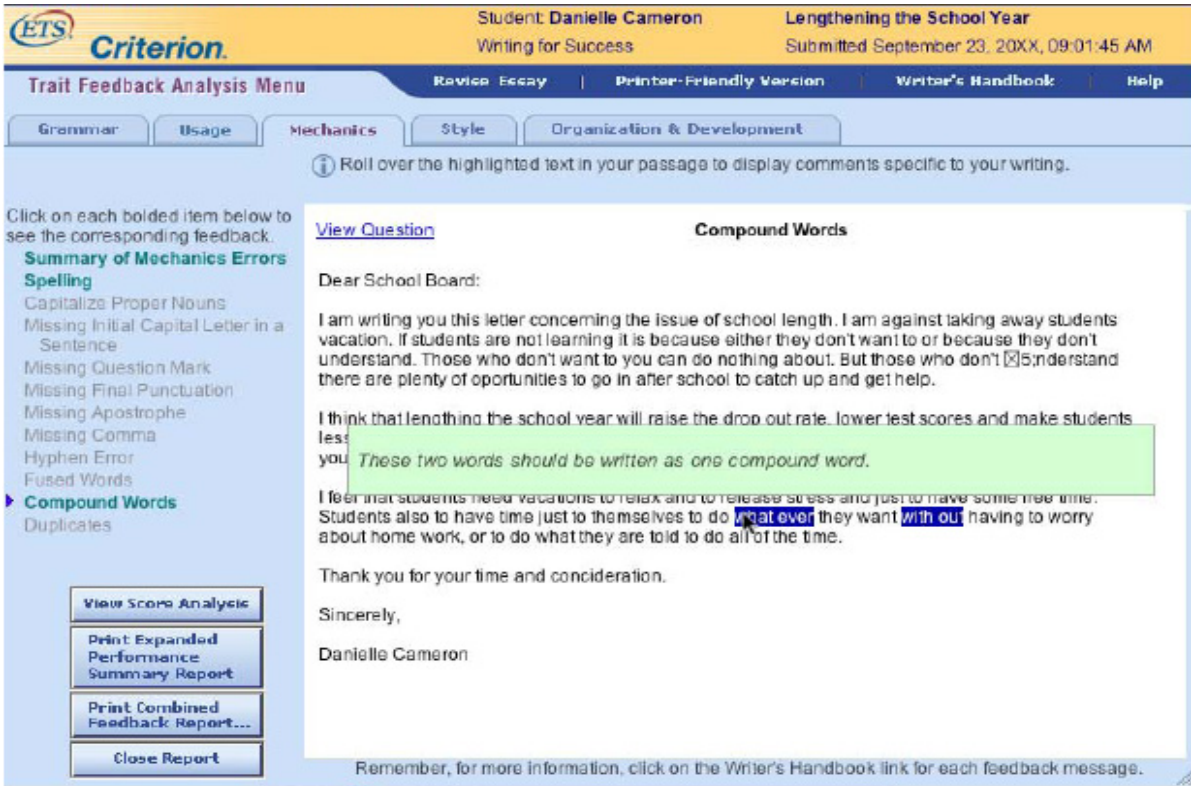


Figure 2.13. Example of verification feedback provided by Criterion

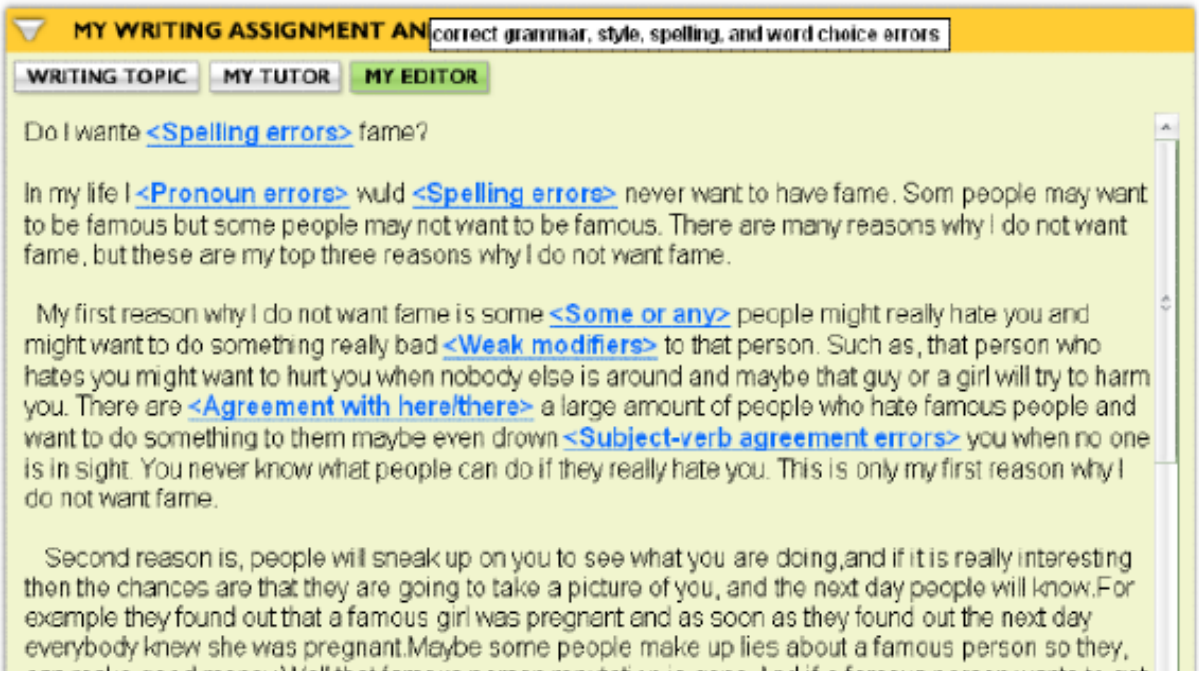


Figure 2.14. Example of elaboration feedback provided by MyAccess!

In writing contexts, teachers rarely have time to provide specific feedback. Their feedback is oftentimes cryptic, vague, and unclear, taking the form of such short questions and comments on the margins as "Why?, Relevance?, Explain!" (Ferris, 2003, p. 26), which make the students frustrated and even resentful (Ferris, 1995, Straub, 1997). AWE systems, on the other hand, rely on the functionality of syntactic, morphological, and discourse parsers as well as on a number of lexical similarity measures (e.g., latent semantic analysis and word sense disambiguation) to analyze, evaluate, and provide specific feedback on various aspects of writing quality such as grammar, style, mechanics, syntactic complexity, topical content, rhetorical structure, and even plagiarism.

Deliberations of how the qualities of effective formative feedback can be reflected in AWE programs acquire a more solid ground if automated feedback is directly compared to other sources of feedback possibly offered in academic writing classes. Conventional CALL feedback is the least advantageous, for CALL applications are not capable of analyzing written constructed responses. Teacher feedback, on the other hand, is extremely valuable, but there are certain limitations to human ability as well. As summarized in Table 2.4, AWE feedback holds substantial advantages in enhancing all the formative feedback qualities.

Table 2.4. Formative feedback qualities in teacher and AWE feedback

Formative feedback quality	*Teacher feedback*	*AWE feedback*
Cognitive involvement	- Points to 'gaps', but possibly inconsistently or subjectively - Not available for scaffolding during writing	- Points to 'gaps' objectively and reliably - Available for scaffolding during writing - Allows monitoring the quality during writing
Goal-orientation	- Not available at any point during writing	- Available at any point during writing

Table 2.4. Formative feedback qualities in teacher and AWE feedback (continued)

Formative feedback quality	*Teacher feedback*	*AWE feedback*
Goal-orientation	- Hardly goal-oriented	- Offers opportunities to self-assess progress - Goal-oriented
Complexity	- Tends to be more verifying and less elaborating	- Informative-tutoring, combining verification and elaboration
Specificity	- Tends to be of selective specificity - Not always individualized	- Very specific - Individualized
Timing	- Mostly delayed	- Immediate - Possibly considering student level and task difficulty - On demand, if needed

Automated writing evaluation and feedback appears to have several important advantages for the formative assessment aspect of L2 academic writing. First of all, a number of practical obstacles would be greatly reduced once the requisite programming is in place. AWE programs can inexhaustibly provide feedback in response to student work, fulfilling the key objectives of formative assessment. Unlike teacher comments, automated feedback can remain unbiased, accurate, and nonjudgmental. It can also potentially increase the students' role in the learning process, enhancing their cognitive engagement and goal orientation due to its timely, complex, and at the same time detailed nature. In addition, AWE programs can be designed to include not only natural language processing techniques employed for the analysis of student input, but also specific databases or student models used to monitor the provision of feedback. Thus, automated feedback can, at least theoretically, be adapted to the level and needs of each individual student, a goal that is not easily attained in traditional classrooms.

2.4.3. AWE issues

Due to the increasing demand for educational technology and to the publicity of vendors, AWE programs are implemented or at least considered for use in a growing number of institutions, e.g., elementary, middle and high schools, community colleges, universities in the United States and abroad, national job training programs, and military institutions (Burstein, Chodorow, & Leacock, 2004). AWE is perceived as a perfect solution by some and as a threat by others.

The supporters of AWE use in the classroom argue that the immense advantages of these programs are their ability to assess and respond to student writing as well as humans do (Attali & Burstein, 2006; Pearson Education, 2007; Vantage Learning, 2007), only in a much more time and cost-effective way. AWE is believed to motivate and guide student revision and to foster learner autonomy. The feedback provided by AWE programs is meant to support process writing approaches that emphasize the value of multiple drafting through scaffolding suggestions and explanations. Their integration into the curriculum is said to also be consistent with the drive toward individualized assessment and instruction. The developers of these programs promote them as instructional supplements to process writing instruction and as vehicles of consistent writing and evaluation across the curriculum.

The truth of such publicity is questioned by members of the academic community, who are skeptical, especially when it comes to the consequences of AWE classroom implementations (Ericsson & Haswell, 2006). Cheville (2004), for instance, takes a very critical stance towards AWE. She is concerned that “early acculturation to such a program might undermine the language and learning of students” (p. 48) and that the machine

analyzer is calibrated to static compositional features and formulaic expression, heavily subordinating meaning. Therefore, Cheville suspects, AWE use may encourage students to pay more attention to the surface features that are more easily detected by AWE systems than to the construction of meaning for communicative purposes. She even fears that "automated scoring technologies make it possible to eliminate the evaluative influence of teachers altogether", and that "what we know and what we do in the classroom is in jeopardy" (p. 49). Her prognosis is that "assessments produced by private corporations will dictate how and what our students know" (p.51). Another worry is that, although it is unlikely that students trick AWE software, they may consciously or unconsciously adjust their writing to meet the assessment criteria of the software. Moreover, teachers may feel pressured to support such adjustment in an attempt to raise test scores (CCCC, 2006). This negative influence, consequently, might undermine teachers' theoretical and practical knowledge.

In addition, questions are raised from the theoretical point of view. It is contended that the social and communicative dimensions of writing are not supported in AWE systems since they are grounded in a cognitive information-processing model (Ericsson, 2006). Student essays are evaluated automatically against generic writing traits, eliminating the value of human audiences in real-world contexts. "While they [AWE programs] may promise consistency, they distort the very nature of writing as a complex and context-rich interaction between people" (CCCC, 2006).

The arguments against the implementation of AWE programs in instruction, overall, express significant concerns that are related to washback, defined as the influence of assessment on teaching and learning (Shohamy, Donitsa-Schmidt, & Ferman, 1996) and as

"the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning" (Messick, 1996, p. 241). However, because assessment (be it summative or formative) is never context-free, washback is inter-related with ecology, a concept used to describe phenomena in their context and to understand both the context and the interactions that create that context" (Garner & Borg, 2005, p.121). The dispute between AWE supporters and opponents is whether this technology causes positive or negative washback, but not much consideration is given to context related factors that may cause one or another kind of washback.

2.4.4. AWE validation

AWE programs have largely been evaluated and validated through research carried out by the commercial companies that have produced these products. Therefore, Warschauer and Ware (2006) warn that "research conducted to date should be considered with a highly critical eye" (p. 7). Indeed, as it will further be demonstrated, while the psychometric analyses results may be strong, there is much more to be considered about the validity of AWE when it is implemented in classroom contexts.

Validity and reliability are related in that the latter is a necessary prerequisite attribute for valid interpretations of assessment outcomes. In automated scoring, the same measure is generally used to evaluate both reliability and validity, and that is agreement between human-computer scores and scores assigned by various numbers of human raters. Correlations and agreement rates between the system and human assessors are typically high, cross-validated correlations between the scoring engines generally ranging from .70 to .90 and often being in

the .80 - .85 range (Dikli, 2006). Automated scores also behave relatively well when compared with external measures such as standardized tests and a range of non-test measures (Keith, 2003).

Although the psychometric results are very encouraging, evincing that reliability is not an issue since automated scoring algorithms yield consistent scoring results, the evaluation of AWE programs should involve more complex analyses to include not only validation of the scoring system independent of the context of application, but also validation of the scoring system used in the application context (Chung & Baker, 2003, p. 27).

The main question that teachers and administrators who intend to implement AWE have is about the impact that AWE programs have on learning, or, in other words, they want to know whether these programs help improve student writing. Few studies have been conducted to address this question, mostly in L1 writing contexts, and their results differ, to say the least. Because of the paucity of research on AWE in ESL/EFL contexts, further discussion includes general research, noting the cases where AWE programs were used with language learners.

Elliot and Mikulas (2004) reported that student writing skills, as measured by performance on statewide writing assessments, were significantly improved by using *MyAccess!*. Their survey results indicated that students were highly satisfied with the automated feedback on their essays and judged it as both helpful and accurate. On a broader range, school districts' experience in the USA shows that the application of *MyAccess!* seems to lead to an increase of writing proficiency throughout the school year and in performance on state assessments as well as to an improvement in the quality of student writing (Vantage

Learning, 2007). These claims are anecdotal rather than empirically supported; they rely on case studies published on Vantage Learning website.

In Foltz, Laham, and Landauer (1999), in an undergraduate course, students used *WriteToLearn* to revise their essays. Out of 100 point, their average first essays score was 85, and the average last essays score was 92. The researchers recorded an improvement in scores ranging from 0 to 33 points over an average of 3 revisions. Attali (2004) investigated how *Criterion* was used nationwide by students in 6th through 12th grade during a school year. The findings demonstrated that automated scores for essays submitted more than once increased from first to last submission, and students significantly lowered the error rate, but the revisions were made mainly at the level of spelling and grammar and not so much at the structural level. Similarly, in Leah Rock's (2007) study, 9th graders used *Criterion* for four weeks, and during this short period of time they received higher analytic scores on their essays written at the end of the study period and improved the mechanical aspects of their writing.

However, when Shermis, Burstein, and Bliss (2004) compared the performance of high school students who were randomly assigned to either a treatment group or a control group, no significant differences were found between the two groups. Warschauer and Grimes (2008) investigated the use of *Criterion* and *MY Access!* in middle, junior high, and high schools. Their participants also showed no significant improvement, and their revisions mainly focused on spelling, punctuation, and grammatical errors and much less on content improvement. On the other hand, they did not find that the AWE software promoted stilted writing or that it distorted the way teachers taught. It is worth mentioning that none of these

studies reported that students adjusted their writing to meet the scoring criteria of the AWE programs.

Although not all these results speak in favor of AWE programs and their feedback, it is too early to draw negative conclusions. Most of these authors indicate in one way or another that the programs were often misused. For instance, some teachers used the AWE program as an explicit form of test preparation (Warschauer & Grimes, 2008). One particular teacher simply adapted the program to her usual non-process writing approach due to its scoring, not feedback capabilities. She was convinced, based on her experience, that "it doesn't matter if they [students] get a lot of feedback" (Warschauer & Grimes, 2008, p.32), thus obviating the need for much revision. Warschauer and Grimes (2008) also uncovered some paradoxes. First, the participating teachers valued revision, but scheduled little time for it. Therefore, it is not surprising that students focused on grammar and mechanics; to revise at more complex levels, students need to become more deeply cognitively engaged, and for that they would need sufficient time. Plus, the limited revision, as the authors explain, is consistent with more general practices in public schools, where revision "invariably focuses on a quick correction pointed by the teacher or peer" (p.29). Second, teachers rarely used AWE in their classes, although they expressed very positive views about those programs. The teachers explained that this was because much of the curriculum was in reading and language arts and not in composition. Another reason was that some writing tasks did not fall within the range of the types of writing that came with the program (e.g., newspaper articles, brochures, letters, etc.)

Considering this, Warschauer and Grimes (2008) argue that "[n]o matter how much teachers claim that they like a type of software […], if they find various reasons not to use the software, it cannot be expected to have much impact" (p.28). Indeed, Shermis et al., (2004) speculated that their discouraging results were partly due to poor implementation and high attrition, with only 112 of the 537 treatment students completing all the essays. They also estimated that if students had written five more writing assignments each, their performance would have significantly improved. This may also be the case in Attali (2004) and Warschauer and Grimes (2008), where the students did not exploit the revision capabilities of the AWE programs; 71% and 72% of the students, respectively, submitted their essay to the system only one time, without revising it.

Further, the effectiveness of AWE may possibly be affected not only by teachers' attitudes, but also by their technological literacy along with student characteristics and level of proficiency (Chen & Cheng, 2008). For instance, according to Warschauer and Grimes (2008), teachers with relatively little computer experience were more reluctant to use these programs in their classes, while those with better computer skills were more enthusiastic about AWE. Unlike higher level of literacy students, students performing below grade level found it difficult to understand the automated feedback other than its most basic aspects, and their teacher's not explaining it did not make the experience easier, which is in line with Chen and Cheng's (2008) finding that implementing *MY Access!* with minimum teacher facilitation can even result in frustration and negatively affect learning. However, Yang (2004) shows that more advanced EFL learners appeared to show less favorable reactions toward the AWE feedback compared to students with lower level of language proficiency.

Finally, Chen and Cheng (2008), who also targeted EFL classrooms, found that the program was perceived more favorably at earlier stages of drafting and revision, especially if teacher and peer feedback were provided at a later point of the writing process.

To sum up, the few existing studies shed little light on AWE washback and on the way it reshapes teaching and learning to write. On the one hand, there is positive evidence that indicates improvement in students' performance; on the other hand, there is negative evidence that points to superficial revisions. Given the inconclusive nature of the empirical evidence, researchers's recommendation that AWE programs be only used as supplements to writing instruction (Shermis & Burstein, 2003; Ware, 2005; Warschauer & Ware, 2006) is more than justifiable. As of now, there is no doubt that the design of AWE software needs to be improved to reflect writing pedagogical theories, but it is also important that more effective pedagogical practices are developed in order to strengthen the positive and attenuate the negative effects that AWE technology may bring to students in their learning to write.

Furthermore, while opponents of AWE argue that its implementation can lead to undesirable consequences, it seems that it is not necessarily AWE that conditions those consequences. As Warschauer and Grimes (2008) put it, the influence of AWE "is mediated by complex relationships among social and educational contexts, teacher and student beliefs, and prior instructional practices" (p. 34). Therefore, blindly accepting or rejecting these programs is not a viable stand (Whithaus, 2006). Both pro and con judgments about the impact of AWE should be based not on speculations and judgmental hypotheses, but rather on empirical evidence that would reveal how these technological innovations are used in diverse contexts and how ecological factors may define AWE use and washback. Thus, more

investigations of AWE implementations in various writing classes are needed "in order to explicate the potential value for teaching and learning as well as the potential harm" (Williamson, 2004, p. 100).

Existing research studies, although informative, have been criticized for their methodological limitations, and mainly for their primary focus on outcomes, excluding the educational process involved. This may be a major reason why their findings are so contradictory. Consequently, Warschauer and Ware (2006) propose a process-product approach to empirical inquiry with a focus on the interaction between AWE use and outcomes. This can provide a more contextualized picture of the effects of automated evaluation and feedback. Studies in this vein are advised to combine both quantitative measures and qualitative insights from observations, think-aloud protocols, interviews, computer screen capturing or video-documentation of student work as well as from longitudinal and ethnographic studies. This dissertation work adheres to this approach, aiming to provide a better understanding of learning processes and of how these processes relate to specific learning outcomes.

2.5. Evaluation approach and research questions

To meet the needs of the given graduate academic writing course at ISU and to account for the lack of automated evaluation of the research article genre, a new NLP-based program, named IADE (Intelligent Academic Discourse Evaluator), was developed taking into account relevant SFL, SAT, and Interactionist SLA theoretical beliefs. This program analyzes students' research article introductions and provides intelligent, discourse-related, and discipline-specific individualized feedback. Considering that the formative value of

AWE in NNS instructional contexts is still an empirical question, as is the value of intelligent feedback, the purpose of the proposed study is to investigate the effectiveness of IADE and the capability of its feedback to enhance the formative assessment aspect of NNS graduate academic writing instruction. Such a goal makes it an evaluation study.

CALL evaluation is generally conducted in two major ways: during and upon the completion of the application's development process. In the first case, experts and users are involved in prototype testing in order to improve its design and functionality (e.g., Blyth & Davis, 2007). In the second case, the final product is evaluated either judgmentally (e.g., Jamieson, Chapelle, & Preiss, 2004) or empirically (e.g., Jamieson et al., 2005; Ma, 2008) in terms of its effectiveness for the intended purpose. A combination of both evaluation types is recommended. Having gone through a considerable prototype testing phase, IADE's evaluation in this study focuses on an empirical evaluation of its effectiveness.

Hubbard (1988) argued that CALL evaluation should be directed by a framework, i.e., a "tool through which an evaluator can create his or her own questions or develop some other evaluation scheme" (p. 52). It is now commonly agreed that a comparative framework for empirical evaluation, in which experimental or quasi-experimental design studies contrast learning outcomes of CALL and non-CALL activities, has inherent biases and limitations in that it greatly reduces the role of context (Garrett, 1991; Chapelle, Jamieson, & Park, 1996; Jamieson et al., 2005). Since it is the context that may be influencing achievements or failures, CALL must be evaluated "in terms of context-specific arguments supported by rationales and evidence based on theory and research in instructed second language acquisition" (Jamieson et al., 2005, p. 94). Consequently, Chapelle (2001, 2003) proposed a

more process-oriented approach to the empirical evaluation of CALL, which places an emphasis on identifying evidence in learners' observable behavior during the completion of language learning tasks and focuses on both the materials and the ways in which learners interact with them. Drawing from instructed SLA perspectives, her conceptual framework puts forth six criteria for evaluation: Language Learning Potential, Learner Fit, Meaning Focus, Impact, Authenticity, and Practicality.

These criteria can be of excellent guidance for CALL research of evaluative nature. For IADE, specifically, which was designed to enhance the learning process through its intelligent feedback, four criteria – Language Learning Potential, Learner Fit, Meaning Focus, Impact – are particularly suitable in that its elements resonate with the relevant theoretical points discussed in Section 2.1:

Interactionist SLA: input, feedback, learner-computer interaction, output

Skill Acquisition Theory: practice leading to automatization

Systemic Functional Linguistics: contextual functional meaning, texts as a mean of professional (social) interaction.

Analyses of evidence pertaining to each of the categories can yield a picture of the effectiveness of the program for the intended purpose (i.e., to help the students develop the skill of writing academically in a way that meets genre and disciplinary norms) in the intended context (i.e, graduate NNS academic writing courses). From the perspectives of Language Learning Potential and Meaning Focus, it can be elucidated whether and how IADE's feedback stimulates focus on linguistic forms (here, moves) used to express various contextual meanings. Further, whether the types of feedback and the kind of writing practice

offered by the program are useful for learners in the targeted academic writing courses can be clarified from the Learner Fit point of view. Finally, Impact can guide the examination of whether and how the students benefit from the interaction with the program. In view of these considerations, the following Research Questions were formulated:

- *Language Learning Potential* – What evidence suggests that the feedback provided by IADE leads to students' noticing of and focus on discourse form? What evidence suggests that students learned the target discourse forms that were focused on during the interaction with IADE?
- *Meaning Focus* – What evidence suggests that the program's feedback directs students' attention towards the meaning of the targeted discourse forms?
- *Learner Fit* – What evidence suggests that the type of formative feedback on targeted discourse forms is appropriate for students with the characteristics of the intended learners?
- *Impact* – What evidence suggests that students have a positive learning experience with IADE?

2.6. Chapter summary

This chapter reviewed the literature that is related to and inspired this dissertation work. It provided a theoretical framework, integrating theoretical insights on language and learning, which allows for making informed decisions with respect to designing, implementing, and evaluating AWE software in a specific academic writing instructional setting. Specifically, it explained how Systemic Functional Linguistics views on language use and how views on learning in the Skills Acquisition Theory and Interactionist SLA in are

applied in the current work. Further, connections were made between relevant constructs from this theoretical framework and formative assessment, the discussion of which was expanded to include computerized intelligent feedback. Then, existing AWE systems were analyzed and AWE research was synthesized in order to elucidate important issues in the state of the art of this area. Finally, the conceptual framework and the research questions were introduced. Chapter 3 describes the NNS academic writing context of this study and the tools used in this instructional context.

Chapter 3. Context and tools

Because of the importance of the context for AWE evaluation, this chapter provides a description of the instructional context in which this dissertation research was conducted. It first describes the placement procedure by which international students are placed into the graduate-level academic writing course under consideration. Then, it elaborates on the corpus and genre-based pedagogical approach employed in this course, specifying the disciplines represented in the corpus and providing examples of tasks which are meant to equip the students with declarative knowledge about the writing conventions in their particular disciplines. The fact that the course is typically taught in computer labs allows for the use of computer-based tools such as the Moodle course management system and the WordSearch concordancing program, which are described here as well.

The chapter also reports on materials that were used in the study beyond the corpus-based approach employed in this instructional context. First, the annotation of a corpus of research article introductions and its classroom application is explained. Second, the core material used in this research, the IADE automated writing evaluation program developed for this study, is introduced, and its design is accounted for from practical, theoretical, technological, and empirical perspectives. The practical and theoretical aspects are interwoven in that relevant insights are drawn from the Skill Acquisition Theory, Interactionist SLA, Systemic Functional Linguistics, and theoretical perspectives on academic language in order to help satisfy certain practical learning needs. The technological capabilities of IADE are shaped by the Evidence Centered Design. Finally, the empirical background for the design of the program that concludes this section reviews previous

research that informed the IADE's feedback characteristics and covers the preliminary research required for the development of this program.

3.1. Instructional context

3.1.1. Placement

Similar to many other universities in English-speaking countries, Iowa State University is home to more than 2,400 international students from more than 100 countries. These students are accepted by the university if they meet both academic and language requirements. To meet the language requirement, graduate students need to have TOEFL scores that, depending on the graduate program, range between 530 and 600 on a paper-based test, 197 and 273 on a computer-based test, and 79 and 197 on an Internet-based test.

Once admitted to the graduate programs of their choice and upon arriving on campus, international graduate students are required to take the institutional English Placement Test (EPT), which contains three components – reading comprehension, listening comprehension, and writing. The task for the writing component of the test requires students to respond to a prompt, which includes an excerpt from empirically obtained data and requires students to write an academic essay analyzing the given data, reflecting on it, and drawing conclusions based on personal opinions. Figure 3.1 below gives an example of an EPT writing task.

Topic: Leisure activities

Task: People enjoy doing different things in their free time. Recently, a survey asked people from the United States what they looked for in choosing leisure activities. First, briefly summarize the findings described in the table below. Then explain whether you think that

Figure 3.1. Sample EPT writing task

people in your culture would have similar or different preferences for spending leisure time than the people from the U.S. who were surveyed. You do not need to discuss all the items in the table and you may add activities that are not mentioned, if you wish. Explain why you think people choose to spend their free time as they do.

Percentage of participants who indicated the listed item to be "very important to you in your leisure activities"
Spending time with your family – 65%
Relaxing – 58%
Having time alone – 55%
Learning new things – 37%
Helping other people – 32%
Creating things – 29%
Meeting new people – 26%
Exercising – 25%
Attending cultural events – 20%

Figure 3.1. Sample EPT writing task (continued)

Students' responses to the writing prompts are scored by human raters who assign them a score with one of three potential meanings. If two raters disagree about the level of a given essay, it is passed on to the third rater, or even a fourth rater if necessary. There are three levels for international graduate students – pass, 101D, an advanced graduate student writing course, and 101B, a course for lower level graduate and undergraduate students. 'Pass' means that students pass the test and are not required to take any English writing courses. Given the levels of the students placed in these courses, English 101B focuses on grammar issues, language errors, and the like. English 101D, on the other hand, is for students who, on the EPT, demonstrate that they can express themselves with a certain degree of ease, effectively organize their thoughts, operate with varied vocabulary, and make use of complex grammatical structures, but have apparent weaknesses in developing academic discourse.

3.1.2. Course description

International students required to take English 101D are either masters or doctoral students in a wide range of disciplines. Therefore, the pedagogical approach currently employed in this course was driven by the need to address the problem of discipline-specificity in writing (Cortes, 2007), concerns about which were expressed both by academics (e.g., Bhatia, 2002; Lee & Swales, 2006) and by graduate students themselves.

Cortes (2007) provides a detailed description of this course, pointing out its three main elements that she compiled and developed: the Iowa State University Academic Writing (ISUAW) corpus, a set of reading materials, and a concordancing program, WordSearch. For this course, the students are encouraged to collect a small corpus of research articles (20 to 30 articles) that reflects their specific discipline. Alternatively, the students, especially those who are only at the start of their graduate studies and have not yet identified a well-rounded direction, can choose to work with the ISUAW corpus, which contains 1,623 research articles (1,322,089 words) published in top professional journals in 40 different academic disciplines (see Table 3.1). The average length of articles is 814.09 words. Individual corpora for each discipline are separated into sections: Introduction, Methods, Results, Discussion, Conclusions, and Abstract. The course work is focused on analyses of these structural RA components one at a time.

Table 3.1. Academic disciplines represented in the ISUAW corpus

Accounting Aerospace engineering Agricultural and biosystems engineering Agronomy	Forest economy and management Geological and atmospheric sciences Genetics Human Health and Public Performance Immunobiology

Table 3.1. Academic disciplines represented in the ISUAW corpus (continued)

Animal science	Industrial engineering
Applied linguistics	Inorganic chemistry
Architecture	Journalism
Art and design	Mathematics
Biochemistry	Mechanical engineering
Bioinformatics	Meteorology
Biology	Molecular biology
Biophysics	Organic chemistry
Business	Physics and astronomy
Chemical engineering	Plant breeding
Computer engineering	Plant physiology
Computer science	Rhetoric and composition
Economics	Statistics
Engineering	Toxicology
Environmental engineering	Urban regional planning
Food and lodging management	

The course is largely learner-oriented in that the students conduct corpus analyses in order to identify and learn about the writing conventions in their particular disciplines. Cortes (2007) explains that students' class work is organized in a way that combines top-down and bottom-up approaches to corpus analysis. In other words, the students analyze entire texts, at the same time examining the use of lexico-grammatical features at the level of discourse and observing typical pattens of their occurrence. She clarifies that the top-down approach is realized through deductive activities, which are based on the reading materials compiled from genre analysis works, where researchers examine disciplinary writing. The reading materials serve as a reference framework for students, guiding them in their corpus analyses and in the completion of class exercises, an example of which is presented in Figure 3.2.

In the previous class, you analyzed the Methods sections in your corpus with the purpose of identifying the overall structural components of the Methods sections in your discipline. Most likely, many of you found subsections with a content that is equivalent or similar to the ones proposed in class (see below), but under different subheadings.

- Study area
  - Site description
  - Experimental design
  - Data collection
  - Sampling procedure
  - Data analysis
- Subjects
- Procedures
- Instruments/Materials/Equipment

The task for today is to analyze those parts of the Methods sections that can be conventionally categorized as "Study area" components and to answer the following questions:

1. Are there any subsections or parts that are equivalent or similar to the categories listed below? If there are, how are they titled?

- Site description
- Experimental design
- Data collection
- Sampling procedure
- Data analysis

2. Where in the methods section do they appear? Are they separate or blended together?

3. How extensive are they? That is, how much detail do the authors provide in each of them?

4. What kind of content is included in each of the subsections found under the so-called "Study area" in your discipline?

Some of you found that, in your discipline, no part of the Methods section matches the "Study area". However, you identified some, as we call them, "Variations". Your task today, then, is to focus on one aspect of the "Variations" you discovered and answer the same questions above.

Take detailed notes describing the structural and content peculiarities of writing the so-called "Study area" subsections in your field. Do not write in short question-answer form; rather, organize your writing into paragraphs. Write as if your audience is students in your discipline who need to know how to write this part of the RA Methods section. The questions above should help you report on your corpus observations, which will become your guidelines when you write the Methods section for your own research article.

Figure 3.2. Observing structural and content variety of the RA Methods section

The bottom-up approach relies on inductive activities that require the students to observe the linguistic means that are typically used in their discipline. The students are first asked to select words and expressions that are representative of a particular discourse element of a given RA section and that would help them effectively build it (e.g., expressions such as "we first determined…", "was calculated as…", "the proposed approach is mainly demonstrated through…", etc., indicate the research procedure; "the expression of … resulted in…", "the data indicated that…", "differences were found in…", etc., report the findings; "this suggests that…", "it seems noteworthy that…", "could be ascribed to…", etc., are used in the discussion of the results). For the purpose of clarity, they are advised to organize their findings into lists of words/expressions with subheadings according to structural components identified. Then, they use the WordSearch concordancer to observe the linguistic environment in which the selected words/expressions occur. WordSerach is described later in this section.

3.2. Computer-based tools integrated in instruction

3.2.1. Moodle

Following Cortes' (2007) corpus-based design, the course is held in computer labs. The ISUAW corpus is available to students in the lab classfiles, and the reading materials as well as the tasks are provided to students through the Moodle course management system (see Figure 3.3 for a screenshot of the main page). The students upload completed tasks to Moodle forums so that they can access each other's postings. With the help of the forum posts, examples of which are shown in Figures 3.4a and 3.4b, the students are not only able to document the writing patterns that they have observed in their corpora, but to also compare

their corpus findings with those in the same or other disciplines posted by their colleagues. Typically, specific examples extracted from the corpora are included in the forum posts, and these are particularly helpful for those students who are not entirely sure whether they properly understood what characterizes a particular discourse element of the research article. Thus, the forums contribute to students' learning, adding an additional source of reference - peers. At the end of classwork on an individual RA section, the students are assigned to write and submit through Moodle a draft of that section, on which they receive teacher feedback and which would become a part of their research paper to be submitted at the end of the semester.

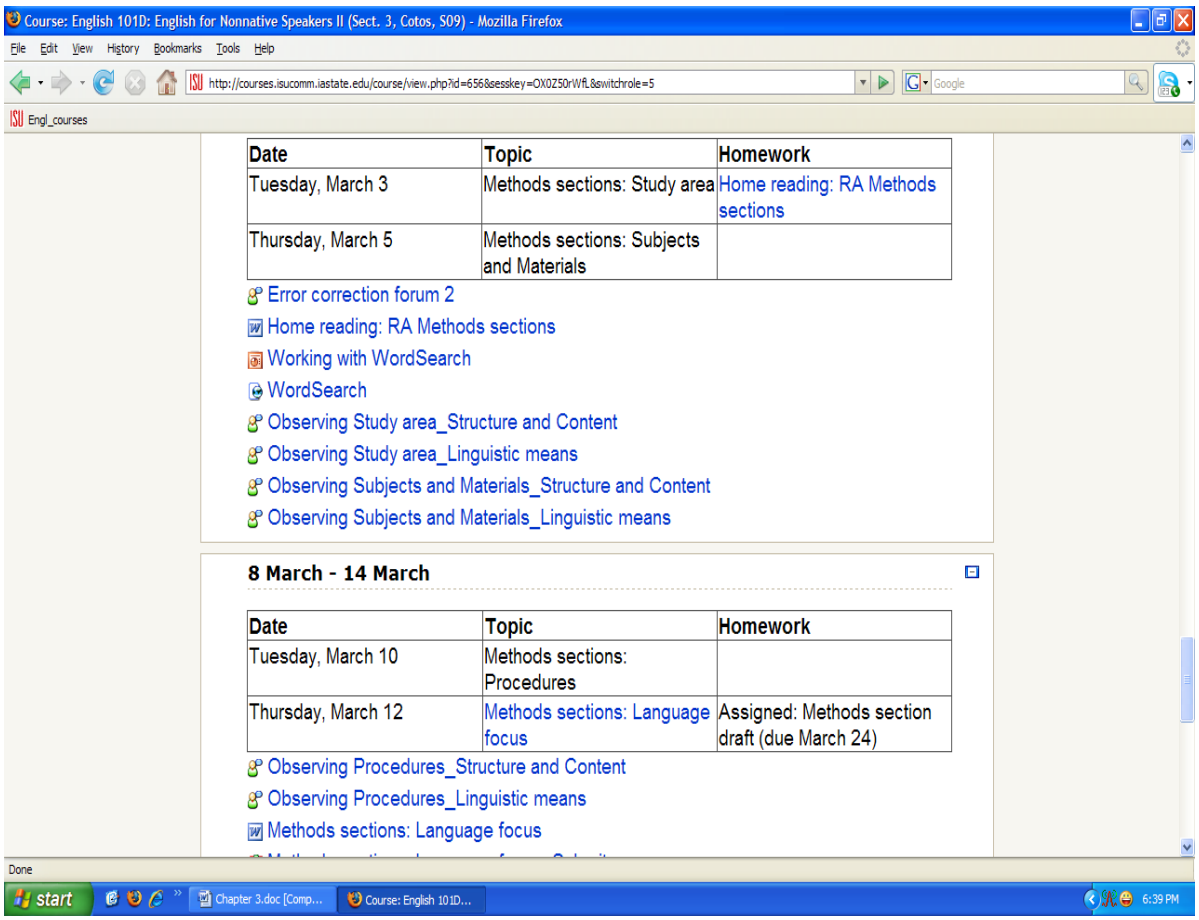


Figure 3.3. A screenshot of the main page in the Moodle course management system

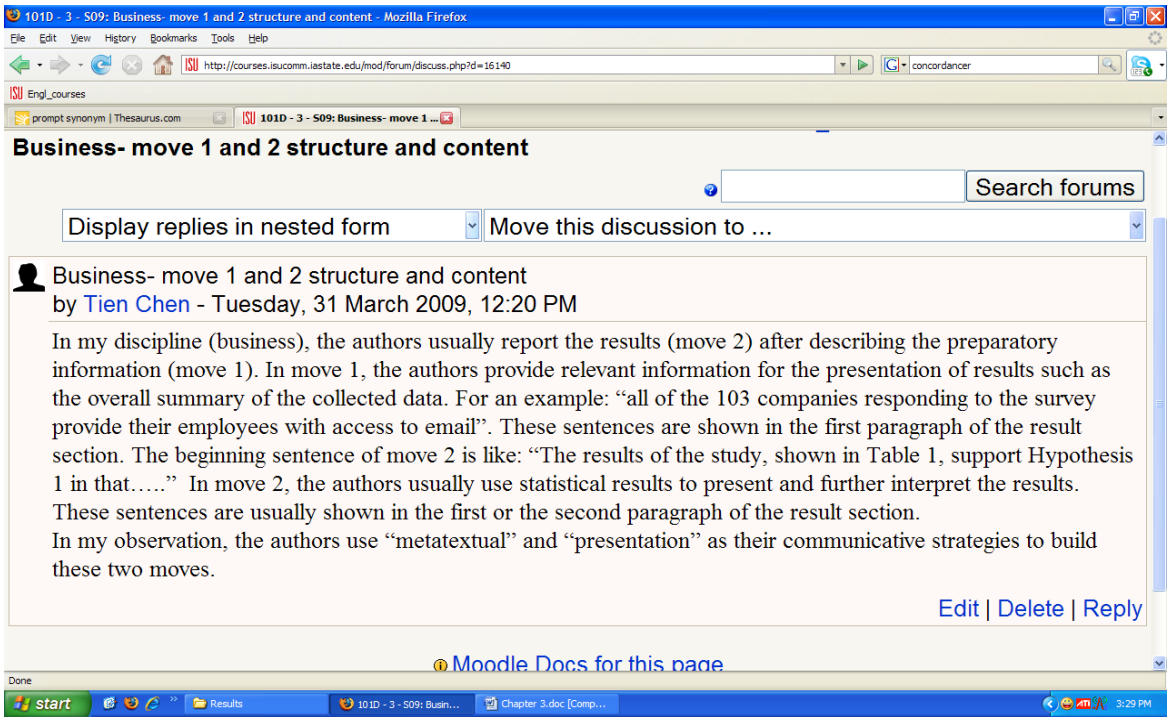


Figure 3.4a. Example of a forum post on research article Results sections' Moves 1 and 2 (structure and content)

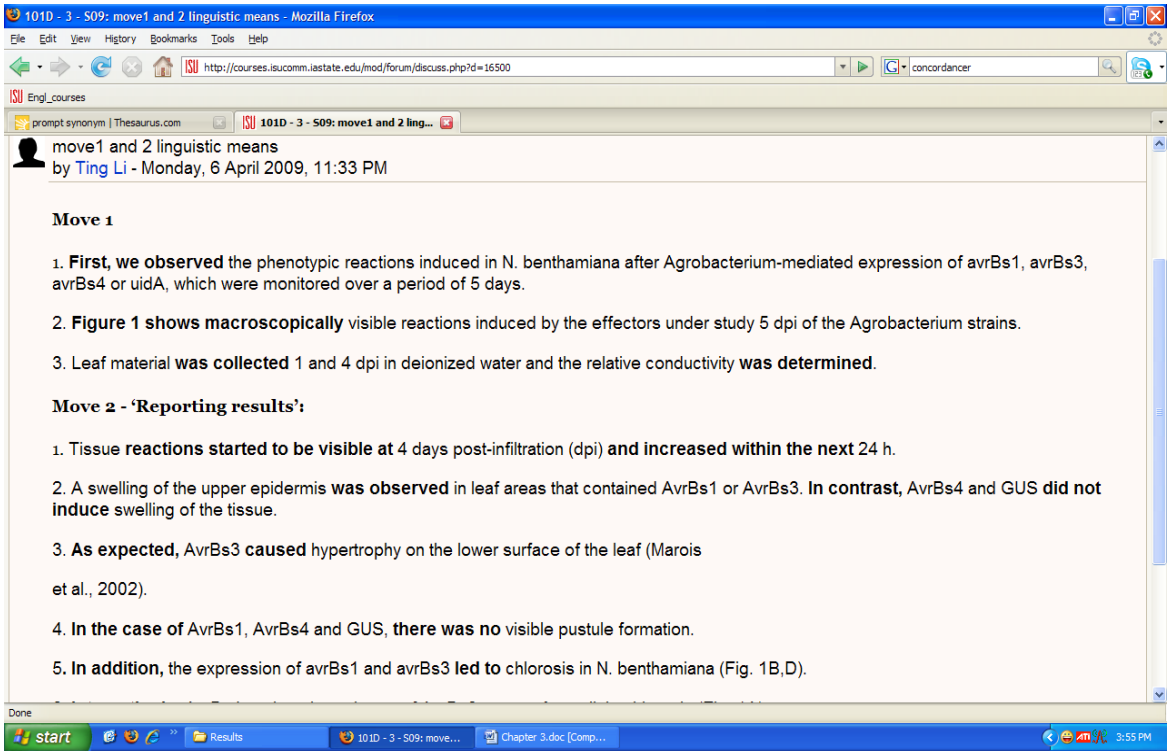


Figure 3.4b. Example of a forum post on research article Results sections' Moves 1 and 2 (linguistic means)

3.2.2. WordSearch

WordSearch allows students to search the corpus for up to four consecutive words as shown in Figure 3.5. The interface of this concordancer indicates the number of matches that result from the query as well as the specific corpus files that were processed. The query output contains stretches of text with 30 words before and 30 words after the searched item; this provides the students with an immediate context in which the linguistic feature was used. The searched items can be words or word combinations that are indicative of certain rhetorical moves. For instance, the 'Announcing present research purposefully' step of move 3 in research article introductions can be marked by "The purpose of this". Figure 3.6 demonstrates and example of such output, where the number at the beginning of each concordance excerpt represents the document from which it was extracted.

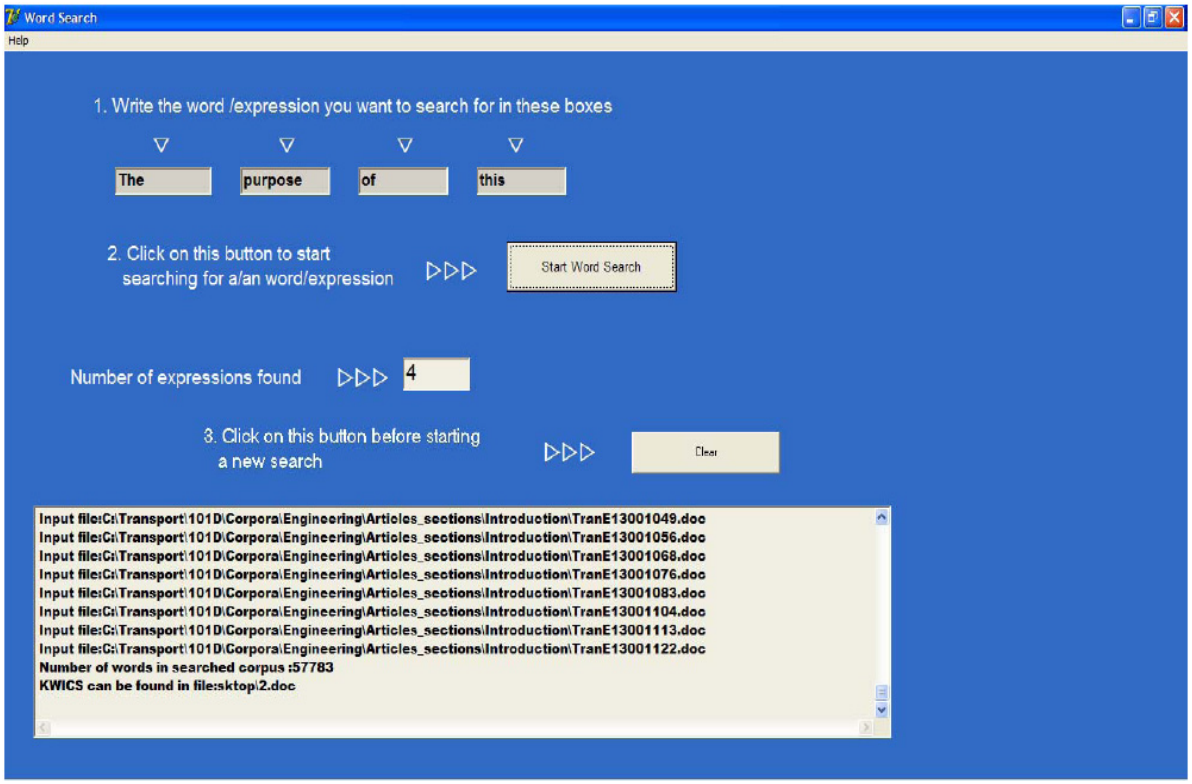


Figure 3.5. Sample WordSearch query

9111010.doc
stress as a result of reduction in pore pressure within the reservoir due to hydrocarbon extraction, reservoirs generally being ``overpressured'' in that the pore pressure is greater than hydrostatic.

The purpose of this

paper is to review some of the laboratory test data generated by the petroleum industry during the development of these fields within the context of a critical state

9111028.doc
the neutron probe access tubes ~10 m!. The correlation was improved by using monthly calibrations, and a residual standard deviation of 0.019 was obtained for the 1.5-m soil profile.

The purpose of this

study was to evaluate the use of EM induction to monitor average water content and changes in water content in the upper 1.5 m of the soil profile

1502157.doc
old and new binder is achieved through a combination of mechanical mixing and mixing by diffusion. Furthermore, phase stability is required for the binder to remain homogeneous after mixing.

The purpose of this

paper is to present a new and versatile method for measuring the diffusion of rejuvenators into bitumen based on Fourier transform infrared attenuated total re¯ectance spectroscopy ~FTIR-ATR!, which

2903223.doc
short intervals ~such as every 1±10 s!, it is important that the transportation engineering community develop an understanding of the in¯uence of the time interval used in ¯ow measurements.

The purpose of this

research is to quantify the impact of the time interval used in freeway traffic ¯ow measurements ~which will be referred to as the measurement interval!. Following a section

Figure 3.6. Sample WordSearch output

Searches can also include ‘wildcards’, which, when accompanying lemmas, yield instances of both base forms and derivatives. For example, when the wildcard “*” is used with “importan*”, the search returns occurrences of “important”, “importance”, “importantly” (see Figure 3.7). The output of each search is saved in a separate automatically created word or text document.

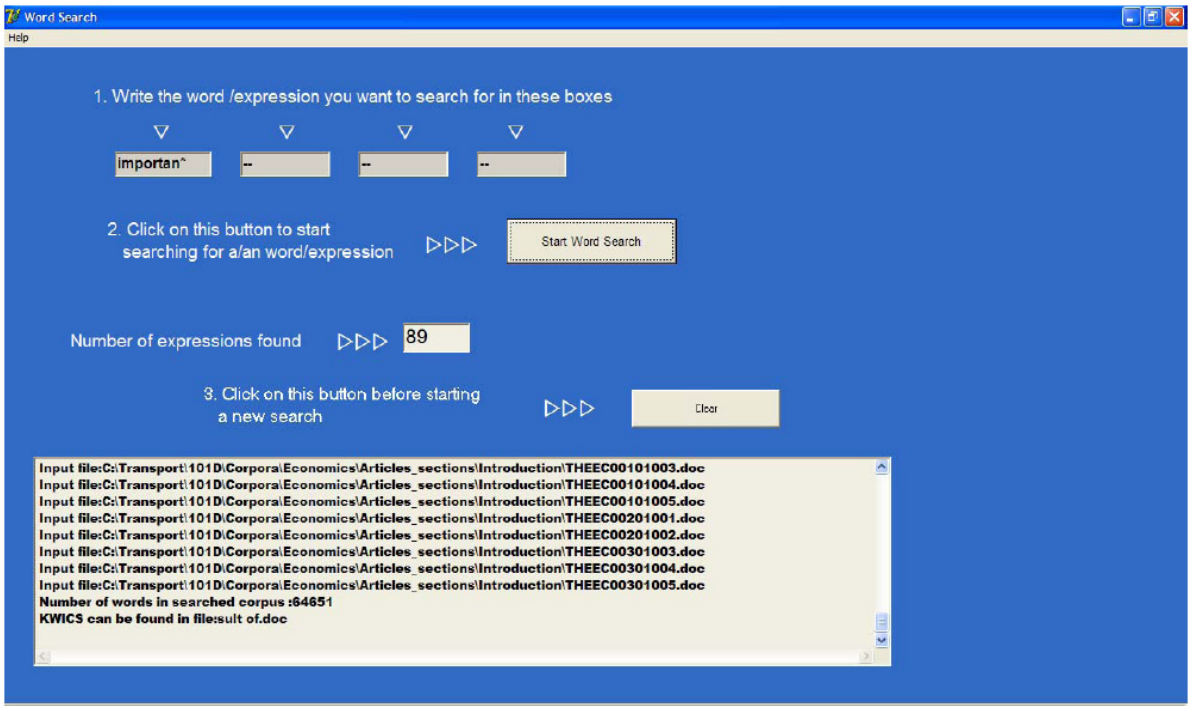


Figure 3.7. A Wordsearch query that includes a wildcard

3. 3. RA Introduction teaching methodology and tools

This dissertation focused only on one part of the English 101D course – the research article introductions. For the purpose of the research, this section of the course has been developed beyond the corpus-based approach in that instruction here uses additional materials and tools, which are described below.

3.3.1. Annotated corpus of RA Introductions

Swales' (1981, 1990, 2004) seminal work on the academic discourse of RA Introductions is used in the teaching of this section. In this part of the course work, the overall corpus-based pedagogical approach described above is enhanced by the use of additional tools. First, the students are provided with a corpus of 1,000 RA Introduction sections representative of 50 disciplines (40 from the ISUAW corpus and 10 additionally

compiled), which is annotated in terms of Swales move schema (shown in Figure 1). The texts are annotated at sentence level, each sentence being labeled with a move and a step within that move (more on text annotation is provided further in Section 3.3.3.3). Consider the following example from the Chemical Engineering corpus:

> <intro_m1 "Metal complexes containing porphyrin, salen, and phthalocyanin ligands are frequently used in oxidation and hydroxylation reactions as chemical models to mimic enzymes [1], [2], [3] and [4]." step= review>
>
> <intro_m1 "A recent trend of research in biomimetic study is to immobilize these organometallic complexes onto a solid support." step= centrality>
>
> <intro_m2 "It is anticipated that this immobilization process will make the catalyst easier to handle especially during separation from the liquid medium." step= hypothesis>
>
> <intro_m2 "Besides, the support might also play certain roles in stabilizing the catalytic activity and/or modifying the chemoselectivity, regioselectivity, and shape selectivity of the catalyst." step= hypothesis>
>
> <intro_m3 "Our long-term research interest is to mimic enzymes, especially hemocyanin which acts as an O2 carrier in mollusk and arthropod." step= purpose> (from CHEE010)

The notation <intro_m1> refers to Move 1, 'Establishing a territory', <intro_m2> refers to Move 2, 'Establishing a niche', and <intro_m3> refers to Move 3, 'Occupying the niche'. As shown in the examples above, the move notation precedes each sentence. Following the sentence is the notation for the step within the identified move. The notations for steps are labeled with a key word from the name of each step, e.g., <step= review> stands for 'Reviewing previous research', <step= centrality> stands for 'Claiming centrality', <step= hypothesis> stands for 'Hypothesizing', and <step= purpose> stands for 'Announcing present research purposefully'.

Annotation is multi-layered for cases when the same sentence signifies more than one move or more than one step. This makes it possible to capture an array of semantic shades rendered by a given sentence. For example, the sentence below combines several steps of Move 3 as is annotated accordingly. "*This paper presents an application of*" introduces the present research descriptively, "*simulation, multivariate statistics, and simulation metamodels*" are indicative of the methodology employed in the study, and "*to analyze throughput of multiproduct batch chemical plants*" states the research purpose.

> "This paper presents an application of simulation, multivariate statistics, and simulation metamodels to analyze throughput of multiproduct batch chemical plants."
>
> <intro_m3 "This paper presents an application of" step= description>
>
> <intro_m3 "simulation, multivariate statistics, and simulation metamodels" step= method>
>
> <intro_m3 "to analyze throughput of multiproduct batch chemical plants" step= purpose>

3.3.2. Callisto

In class, the students use Callisto, an open source annotation tool developed by MITRE Corporation for the linguistic annotation of texts. Callisto allows for unique tag-set definitions, which means that annotators can create and use their own coding schemes and categories depending on their text analysis needs. This tool, however, is used in class not for annotation, but merely for access to the annotated corpus due to its clear and visually appealing display of texts. The main window in this program's interface displays the annotated text (see Figure 3.8). The move-level annotation is reflected by three colors for each move: blue for Move 1, red for Move 2, and green for Move 3, the colors being set as default during prior corpus annotation. By scrolling over or by clicking on a sentence, the

students are able to see the step(s) that the sentence represents and the multiple layers when there are such; this information is displayed below the text window. In addition to that, the students have direct access to examples of a given move by clicking on one of the tabs named “intro_m1”, “intro_m2”, and “intro_m3”. Under these tabs, the names of the steps are displayed next to sentences. A click on a sentence under a move tab highlights that sentence in the main text area, which allows the students to see the contextual use of the given sentence. The annotated corpus accessed through Callisto facilitates students’ corpus analyses in that they are able to see the color-coded overall rhetorical organization of up to 20 introductions in their field. It also exposes them to numerous examples of moves and steps.

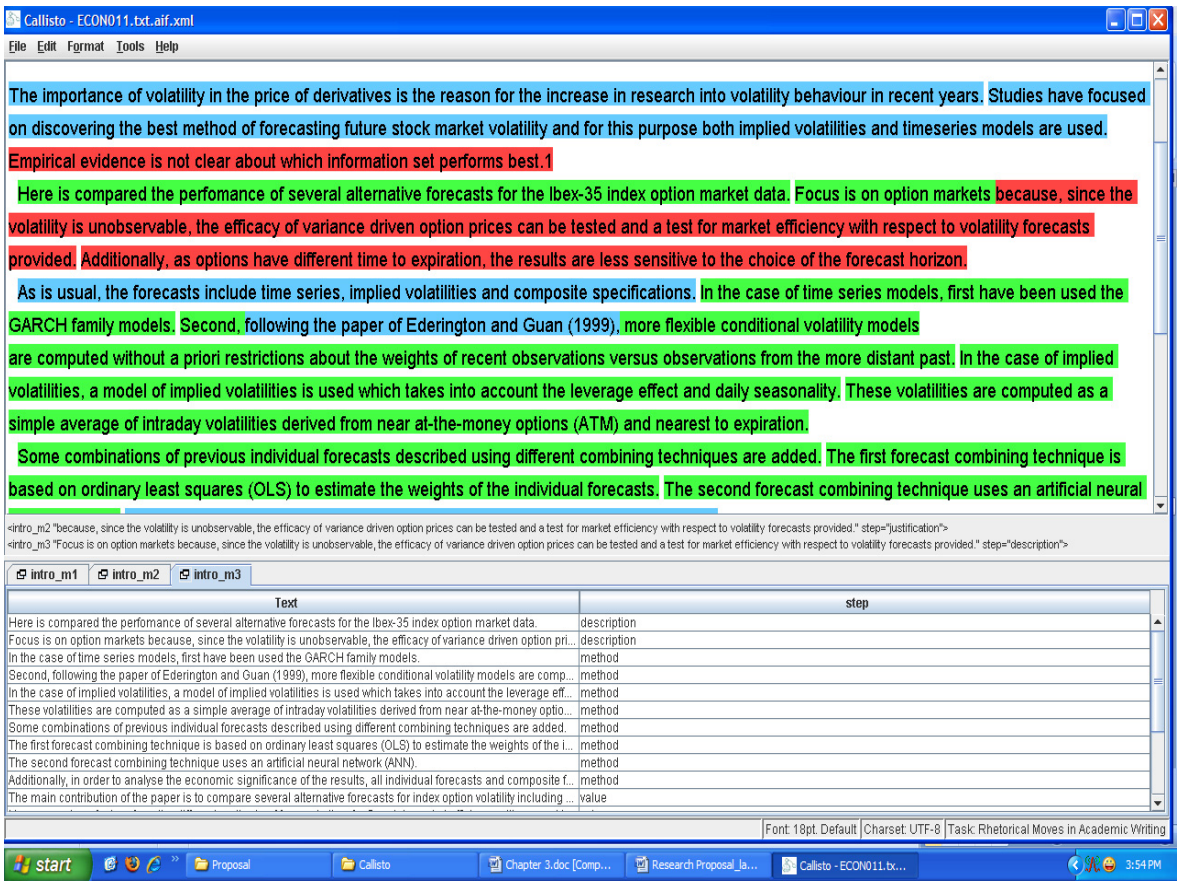


Figure 3.8. Example of an annotated RA Introduction opened in Callisto

3.3.3. IADE

As described in the sections above, in the given writing course, students develop or are provided with corpora of research articles published in top academic journals of their discipline and then conduct corpus analyses in order to discover organizational patterns characteristic of their particular discipline. The class activities also encourage students' use of the WordSearch concordancing tool to explore the corpus and to discover linguistic patterns characteristic of that particular discipline. For RA Introductions, they use Callisto to work with annotated corpora in their particular disciplines. However, this is work done only at the level of recognition, intended to lead only to raising awareness about the genre. Ultimately, the students have to transfer the new knowledge to writing their own research article. This task is the most challenging in that it requires sufficient practice on the part of the student and individualized guided feedback on the part of the instructor.

Therefore, the preliminary goal of this dissertation was to develop IADE (Intelligent Academic Discourse Evaluator), a new AWE program to be introduced as an additional tool for students to practice with and make incremental improvements on their drafts of research article introductions in order to help them apply the knowledge they acquire in the classroom in their own writing. Figure 3.9 sums up all the perspectives integrated in the design of IADE, showing the connections between the practical needs of the targeted students, theory, research, and implemented design decisions. Since the focus of the course is on the discourse patterns and linguistic conventions of research articles, the program does not detect and provide feedback on students' language errors.

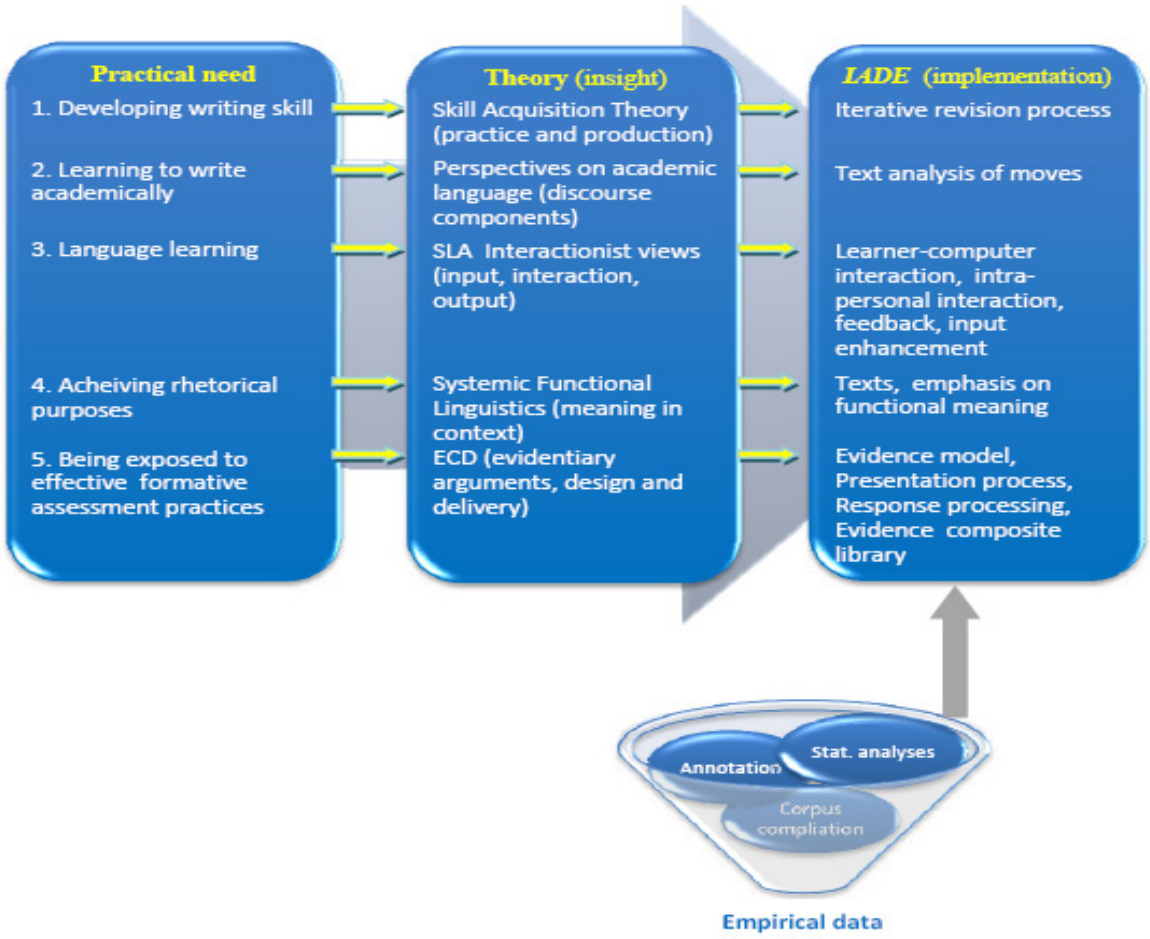


Figure 3.9. Practical, theoretical, and empirical background for the design of IADE

3.3.3.1. Practical and theoretical background for the design of IADE

IADE was developed to meet practical needs. Nevertheless, while practice was the main driving factor, it was informed by a number of other important perspectives. Considerations as to what theoretical insights are most informative vis-à-vis the targeted learners' needs were central to the decisions regarding IADE's design.

First and foremost, the targeted learners need to develop their writing skill. Thus, the stipulation of the Skill Acquisition Theory that repeatedly acting on new knowledge is essential for gradual automatization of that knowledge suggests an iterative revision process,

which the new program makes possible. Second, a closely related need of the targeted learners is to learn how to write academically, abiding by the writing conventions in their fields. Previous discourse analysis research has demonstrated that content and structural development of research articles has its own specifics depending on the academic domain. Indeed, this genre is very complex, a particular degree of complexity being exhibited by Introduction sections, where the writer has to operate with various discourse techniques to fulfill a multitude of purposes. Swales' (1981, 1990, 2004) seminal work on introduction sections offers exceptional guidance, specifying the possible discourse elements, called 'moves'. Therefore, the move schema developed by Swales is not only used in teaching, but also served as a model for the overall approach to the text analysis performed by IADE.

Third, to achieve certain rhetorical purposes in their writing, learners need to know how to express functional meaning. Therefore, as indicated by Systemic Functional Linguistics, IADE focuses on texts with their contextual richness, rather than on sentences with a limited context. It also intends to establish relations between academic texts and the professional practice they realize, and sustains the analysis of production.

Last but not least, since the targeted learners are non-native speakers of English and since writing is viewed in this project as a language skill, it is important to take into consideration what facilitates language learning. According to the Interactionist views on SLA, input, interaction (both human and computer), feedback, and output are viewed as playing a significant role in second language acquisition. Thus, the program analyzes learners' output to generate it into feedback which serves as input. At the same time, it aims at encouraging noticing of and focusing on discourse form by providing input enhancement

of rhetorical moves in the color-coded feedback. The feedback, in turn, is meant to foster intra-personal interaction that is expected to lead to the creation of revised output.

3.3.3.2. Technological background for the design of IADE

The conceptualization and operationalization of the process of feedback generation was guided by the principles of Evidence Centered Design (see Section 2.3.2). The main concept informing IADE is the ECD's Evidence Model central concept of observable evidence, that is, evidence that describes learners' performance as reflected in their work products. From the delivery framework, considering that IADE provides one major task (to revise) and that it is meant for formative evaluation, only the Presentation Process and the Response Processing seemed pertinent and were thus operationalized. Finally, based on the Evidence Composite Library in the ECD framework, a database where all the data are contained was built into IADE. The ECD influence is further reflected in the system overview (Figure 3.10) and the description below. In short, guided by ECD, the program was built to identify the discourse elements of students' work products that constitute evidence and to characterize the strength of this evidence about targeted writing proficiencies by generating automated formative feedback.

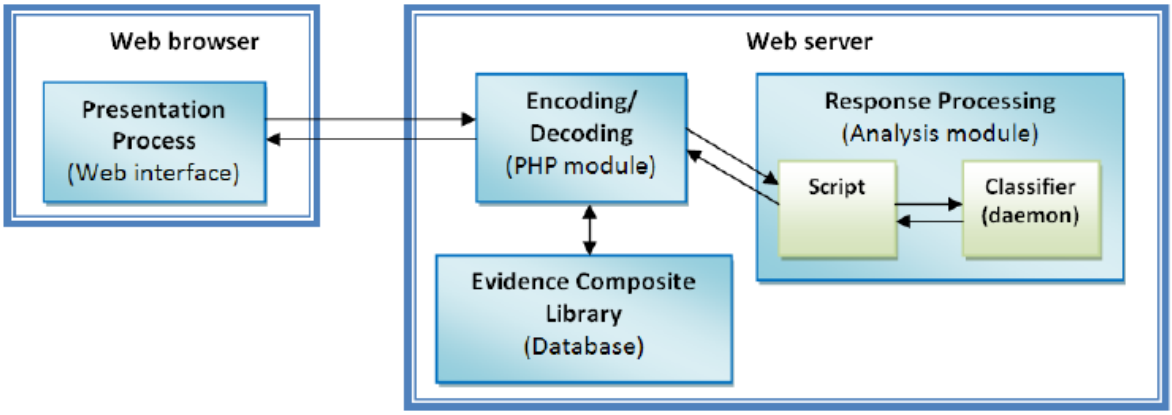


Figure 3.10. IADE system overview based on Evidence Centered Design

The Presentation Process was operationalized through a web-interface where users can access a sign-up page, a log-in page, and a drafting page with instructions, a resubmit button, 'Help Options', and sign-out option. After the first submission, the drafting page displays both the system's feedback (color-coded and numerical) and the revision textbox in which learners make the necessary changes to be submitted for further evaluation.

The Encoding/Decoding Module, which is a PHP (Hypertext Preprocessor) module with a number of extensions, handles queries with the help of the database. It also receives student texts and transforms them into a format recognizable for processing. Then, it gets back the analyzed texts and the numerical feedback from the Analysis Module, transforms the texts in color-coded readable HTML format, and makes all that available on the surface web interface.

The Response Processing is an analysis module, which incorporates two components: a Python script and a Support Vector Machine (SVM) classifier (see Pendar & Cotos, 2008). The Python script executes pre-programmed commands such as breaking the text into sentences and sending each sentence to the classifier for automated analysis. The classifier analyzes and classifies each sentence as belonging to a particular move. The classification is done by means of identifying the lexical features that are indicative of a certain move (the feature selection process is explained further in Section 3.3.3.3). Then, also with the help of pre-programmed scripts, percentages for the move distribution in the student's draft are automatically calculated and compared with the distribution of moves in the annotated corpus of his/her academic field. The classification into moves and the numerical information about

the distribution of moves both in the student draft and in the annotated corpus are included in the feedback, which makes it individualized and discipline-specific.

The SVM model is very big, and that presented a technical problem at the development stage. First, loading and unloading the model for every draft submission would not have been practical, and second, loading more than one model at the same time when multiple users submit their drafts for processing would not have been feasible. However, it was imperative that the system allowed concurrent access by multiple users and concurrent analyses by the same model. To address this problem, the SVM classifier was integrated as part of the analysis module running as a background process on the server, which in the Unix world is known as a *daemon*.

Lastly, the Evidence Composite Library is a database containing the information necessary to make IADE functional. Specifically, it stores lists of:

- User information (first and last name, login name, password, e-mail address)
- User classification (graduate MA, graduate PhD, other)
- Disciplines (50)
- Annotated corpora (50 disciplines, 20 RA introductions each = 1000 introductions)
- Students' submitted drafts (automated analysis and numerical feedback)
- Definitions in Help Options accessed by individual students (number of hits)
- Annotated corpora in Help Options accessed by individual students (number of hits)
- Step statistics in Help Options accessed by individual students (number of hits)
- Revision tips in Help Options accessed by individual students (number of hits)
- Steps within each move

- Step statistics for all 50 disciplines (minimum, average, and maximum percentages)

The 'Help Options' is a feature that was added after the program was piloted with a sample of 12 students in July of 2008. It is meant to enhance interaction and input modification and to satisfy potential students' characteristics and learning styles. This feature includes: Definitions, Step Statistics, Annotated Corpus, and Revision Tips. All these options can be accessed on demand in case students need additional guidance during the revision process. Upon clicking on Definitions, the students are provided with explanations and examples of each move and the steps within a particular move (see Figure 3.11).

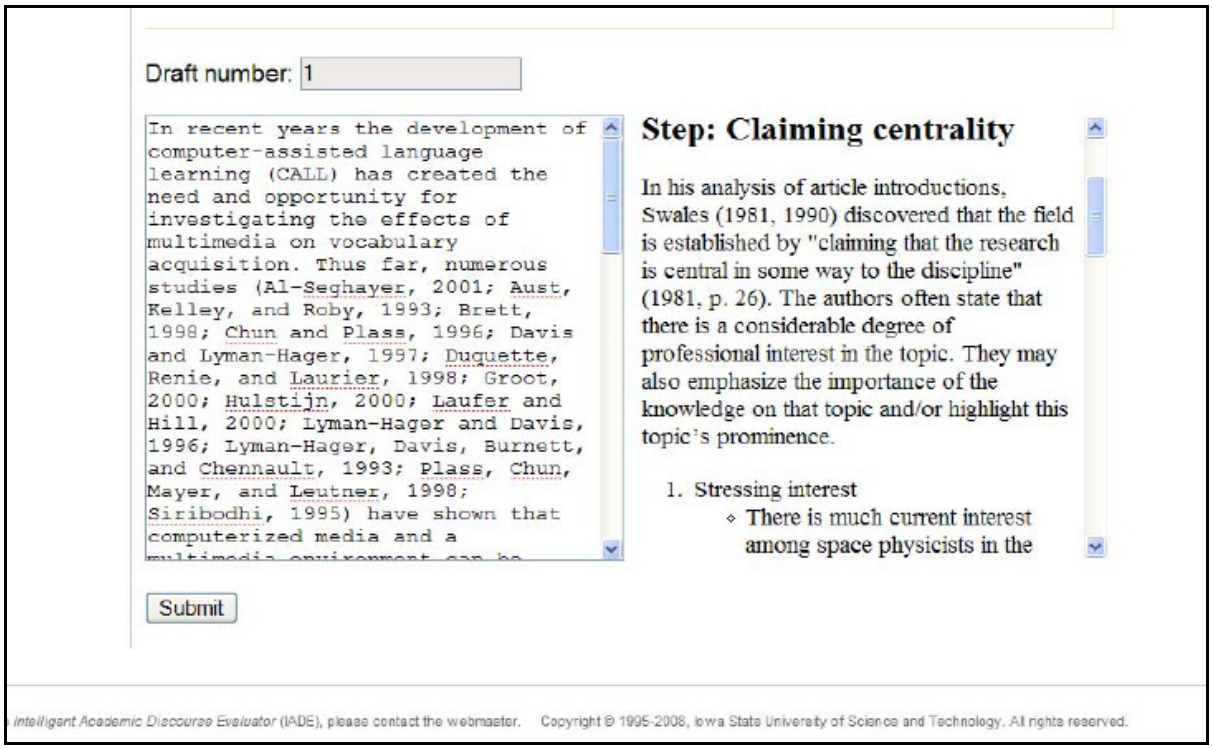


Figure 3.11. Sample step definition and example in IADE's Help Options

The Step Statistics are similar to the numerical feedback returned by the program at the move level, showing the minimum, average, and maximum percentages for each step

distribution in a given discipline. Figure 3.12 shows an example of such information for a student in Computer Engineering.

Steps in Computer Engineering research article introductions

Move 1

Claiming centrality: [min 0%, avg 6.21%, max 21.42%]

Making topic generalization/s: [min 0%, avg 26.71%, max 48.27%]

Reviewing previous research: [min 0%, avg 8.05%, max 36.36%]

Move 2

Indicating a gap: [min 0%, avg 0.91%, max 7.69%]

Highlighting a problem: [min 0%, avg 9.43%, max 36.84%]

Question-raising: [min 0%, avg 0%, max 0%]

Hypothesizing: [min 0%, avg 1.60%, max 13.33%]

Presenting justification: [min 0%, avg 2.56%, max 18.91%]

Move 3

Announcing present research descriptively: [min 0%, avg 16.65%, max 64.28%]

Announcing present research purposefully: [min 0%, avg 1.28%, max 7.14%]

Presenting research questions: [min 0%, avg 0%, max 0%]

Presenting hypotheses: [min 0%, avg 0%, max 0%]

Definitional clarifications: [min 0%, avg 0.56%, max 7.84%]

Summarizing methods: [min 0%, avg 4.54%, max 20%]

Announcing principal outcomes: [min 0%, avg 3.72%, max 13.33%]

Stating the value of the present research: [min 0%, avg 6.34%, max 27.77%]

Outlining the structure of the paper: [min 0%, avg 11.03%, max 33.33%]

Figure 3.12. Sample Step Statistics in IADE's Help Options

When accessing the Annotated Corpus, the students are exposed to 20 research article introductions in their field, which are not only color-coded for moves, but also glossed for steps at sentence level (see Figure 3.13).

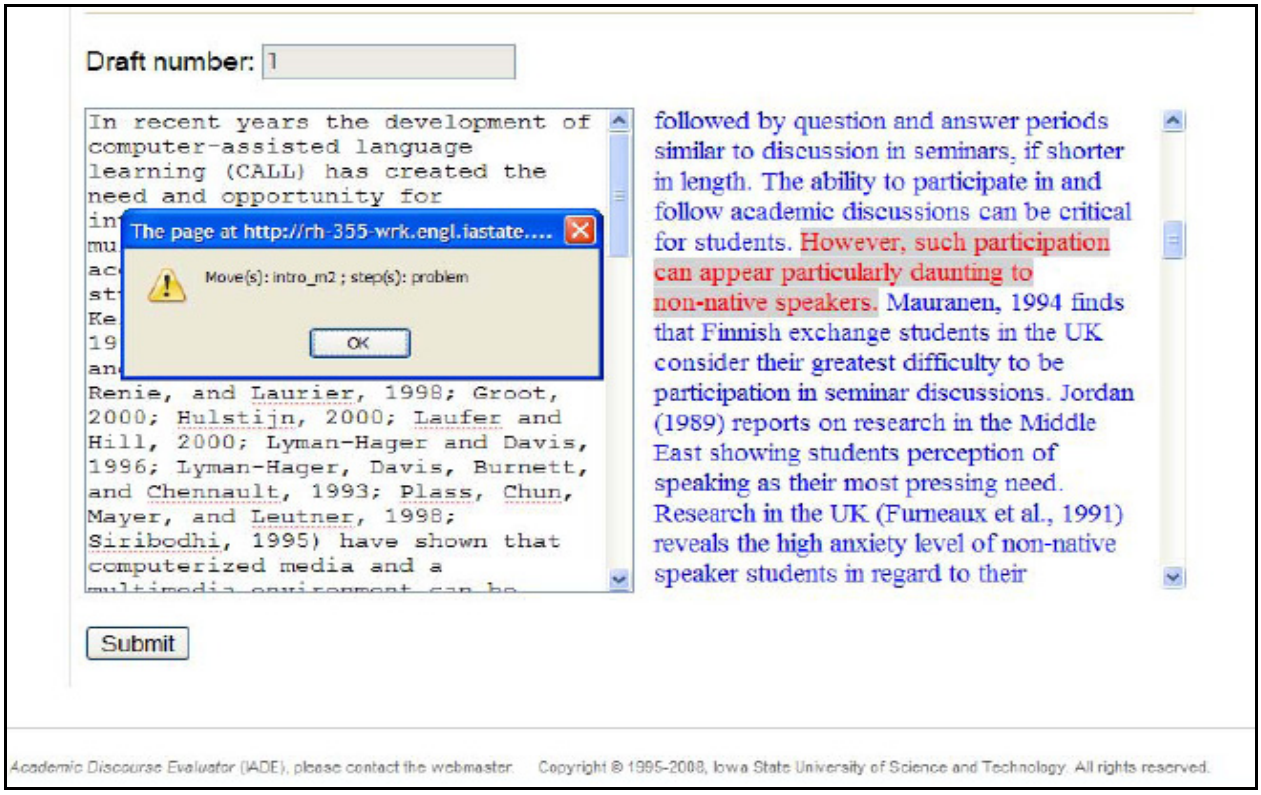


Figure 3.13. Sample annotated text in IADE's Help Options

Finally, the Revision Tips (Figure 3.14) are meant to assist the students in improving their drafts, providing suggestions on what might make their revision process more effective.

Revision Tips

You don't know how to revise or what to begin with? These tips may help you get started and revise your introduction more effectively.

Tip 1. What moves and steps did you have in mind when you wrote your first draft? Take a look at your color-coded text. Are the colors, representing the moves you had in mind shown in the text? If not, you need to make up for the lack of what you are missing and

Figure 3.14. Revision Tips in IADE's Help Options

verify that what the colors you is indeed reflect what you mean.

Tip 2. If the colors representing the moves you had in mind are shown in the text, are they in the order you meant them to be? If not, you might want to consider restructuring the text.

Tip 3. Look at the distribution of your steps compared to their distribution in your discipline, which is shown by percentages. How close are your moves to the 'average' in your discipline? If they, or some of them, are "way above average", "above average", "way below average", or "below average", it means that you need to revise them.

Tip 4. Decide which move you would like to begin revising, but it advisable that you begin with the move that represents the logical beginning of your introduction.

Tip 5. If your structure at move-level seems to be fine, i.e., if all your move are "about average", check your moves for the presence and quality of the steps.

Tip 6. If you need to know how the steps are distributed in the introductions of your discipline and to decide which steps to use, choose the "Steps statistics" in the Help Options.

Tip 7. If you need to see examples of moves, click on the "Annotated Corpus" in the Help Options. There, you will see colors for the moves, and if you click on sentences, a pop-up message will indicate the move and the step within that move.

Tip 8. If you need to see definitions and examples of moves and steps, click on the "Definitions" in the Help Options.

Tip 9. Check if you agree with the color representation of the moves in your draft. Does the analysis reflect the meaning you had in mind? If not, try to identify what may interfere with the program's misunderstanding of your moves and make respective revisions.

Tip 10. Make sure you check if each of your sentences expresses the meaning and the function you had in mind even if you think the colors are correctly displayed by the program.

Tip 11. Be consistent in your revision strategy – if you started revising at move level, then continue revising move by move; if you started revising at step level, continue revising step by step; if you started revising at sentence level, continue revising sentence by sentence, etc.

Tip 12. If you tried to make multiple corrections, and your move/s still appear different from what you mean, this may be because you are not expressing the function of the move clearly enough. Check whether you are using appropriate linguistic means that would signal the intended move.

Tip 13. When revising, you need to focus a lot on the meaning you are trying to convey. Remember, though, that the meaning is effectively expressed not only through words, but also through accuracy of expression and grammar.

Figure 3.14. Revision Tips in IADE's Help Options (continued)

Tip 14. Do not limit your revision by to trying to focus on length only. If your moves are evaluated as "about average", it does not necessarily mean that the introduction is perfect. To ensure that, you need to check whether the meaning of moves is effectively expressed through their steps.

Tip 15. After you make incremental revisions, make sure that the coherence is preserved after the new changes.

Figure 3.14. Revision Tips in IADE's Help Options (continued)

3.3.3.3. Empirical background for the design of IADE

Previous research on feedback

SLA and L2 writing researchers have long debated the issue of feedback (Ferris, 2006). The debates ranged from providing feedback that corrects all errors (Lalande, 1982) to selective feedback (Bates, Lane, Lange, 1993; Ferris, 1995) and even to calls for elimination of any degree of corrective feedback (Cook, 1991; Krashen, 1984; Truscott, 1996). These views have generated an extensive research area. Many works speak in favor of various types of traditional feedback, e.g., explicit (Caroll & Swain, 1993; Lyster, 1998), individual specific (Hyland, 1998), metalinguistic (Rosa & Leow, 2004), negative cognitive (Mitchell & Myles, 1998), detailed iterative (Hyland & Hyland, 2006), and accurate, short, one at a time (Van der Linden, 1993).

Considering the pedagogical and formative assessment intent for the implementation of IADE, decisions regarding its feedback were informed by these research findings as well as by the studies reviewed in Section 2.3.1. To ensure that the program's feedback is meaningful, i.e., a "response that provides a learning opportunity for students" (Heift, 2003, p. 533), it combines a number of characteristics that are beneficial for language learning:

- *Immediate*: the feedback is provided immediately, in less than 60 seconds from the time of submission
- *Intelligent*: the feedback is generated automatically by a natural language processing-based engine
- *Individual specific*: the feedback is provided to the student individually based on his/her input and on its comparison to the respective discipline
- *Metalinguisitc*: the feedback is provided in definitional terms (i.e., "… of your sentences belong to Move 1").
- *Short*: the feedback is concise in that it briefly presents the descriptive percentages representing the distribution of Moves in the students' draft and in the introductions of his/her discipline (e.g., "This is below average compared to Move 1 in your discipline, where the minimum is 45.455%, the average is 65.799%, and the maximum is 87.097%).
- *Negative*: the feedback points to drawbacks in the discourse development of the draft (e.g., the feedback may say: "This is way below average (or below average/ way above average/ above average) compared to Move 1 in your discipline"…)
- *Explicit*: the feedback is clearly stated through comments (way above average, above average, average, below average, way below average) and demonstrated through colors and percentages
- *Output-focused*: the feedback targets the student's actual written production
- *Iterative*: the feedback is provided as often as requested
- *Color-coded*: for input enhancement, the feedback about the draft structure is provided in colors: blue for Move 1, red for Move 2, and green for Move 3

Figures 3.15a and 3.15b contain screenshots of the color-coded and numerical feedback provided by IADE.

Figure 3.15a. Sample color-coded IADE's feedback

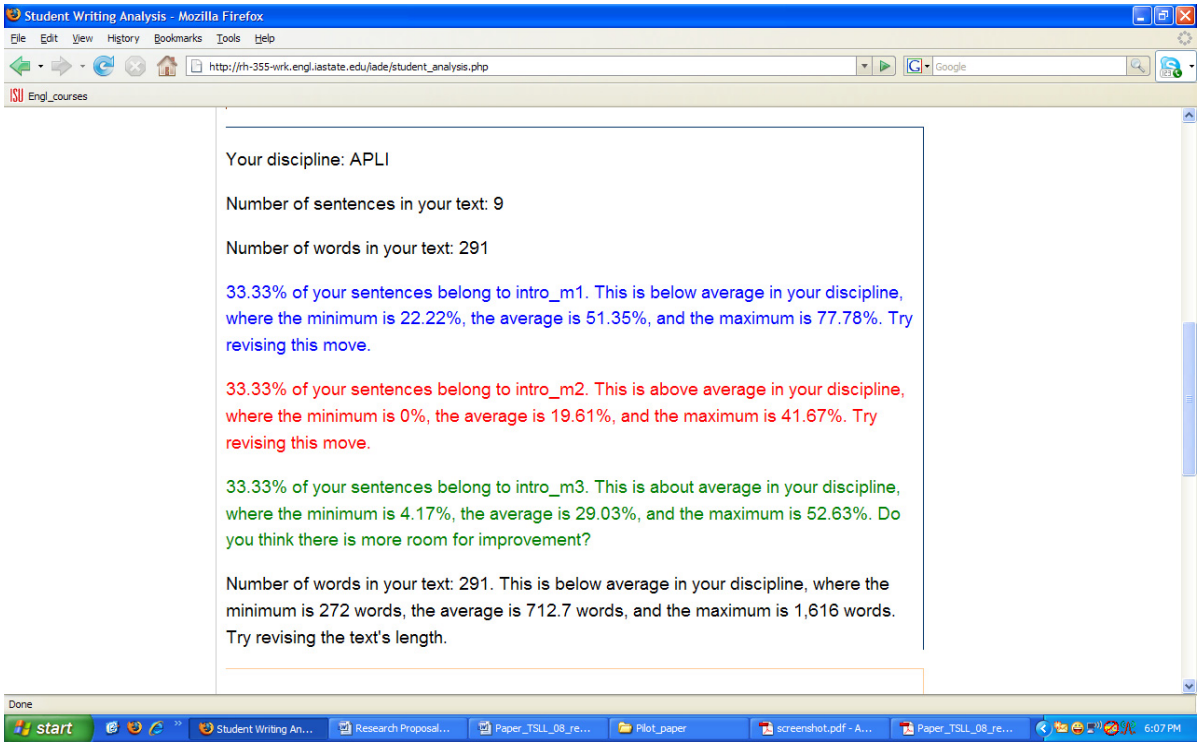


Figure 3.15b. Sample numerical IADE's feedback

Project-specific research on discourse moves

Equally important to the development of IADE was the empirically obtained data that lies at the basis of the program's automated evaluation and feedback. It was obtained through a thorough discourse analysis of a corpus of 1,000 published research article introductions representative of 50 disciplines (the same corpus that is used in the teaching of RA Introductions in English 101D, as described in Section 3.1.3). This corpus of introductions was manually annotated for moves and steps based on Swales' (1981, 1990, 2004) framework.

- RA Introductions corpus annotation

The annotation of 1,000 texts was done in Callisto, the annotation tool introduced in Section 3.1.3. From this corpus, a stratified sampling of 401 introduction sections representative of 20 academic disciplines was made. The size of this sub-corpus is 267,029 words; each file is on average 665.91 words long, resulting in 11,149 sentences as data instances. Figure 3.15 shows an example of text annotation. For annotation, the texts have to be stored as text files. When opened in Callisto, each sentence of the text is selected and assigned a move by clicking on a respective colored tag in the small "Available Actions" box. A step is assigned by choosing a step option from the drop-down menu in the lower part of the screen. The same actions are taken if a second or even third layer is necessary. When text annotation is complete, the new files are saved as .xml documents so that they can be open in Callisto to display the assigned moves and steps.

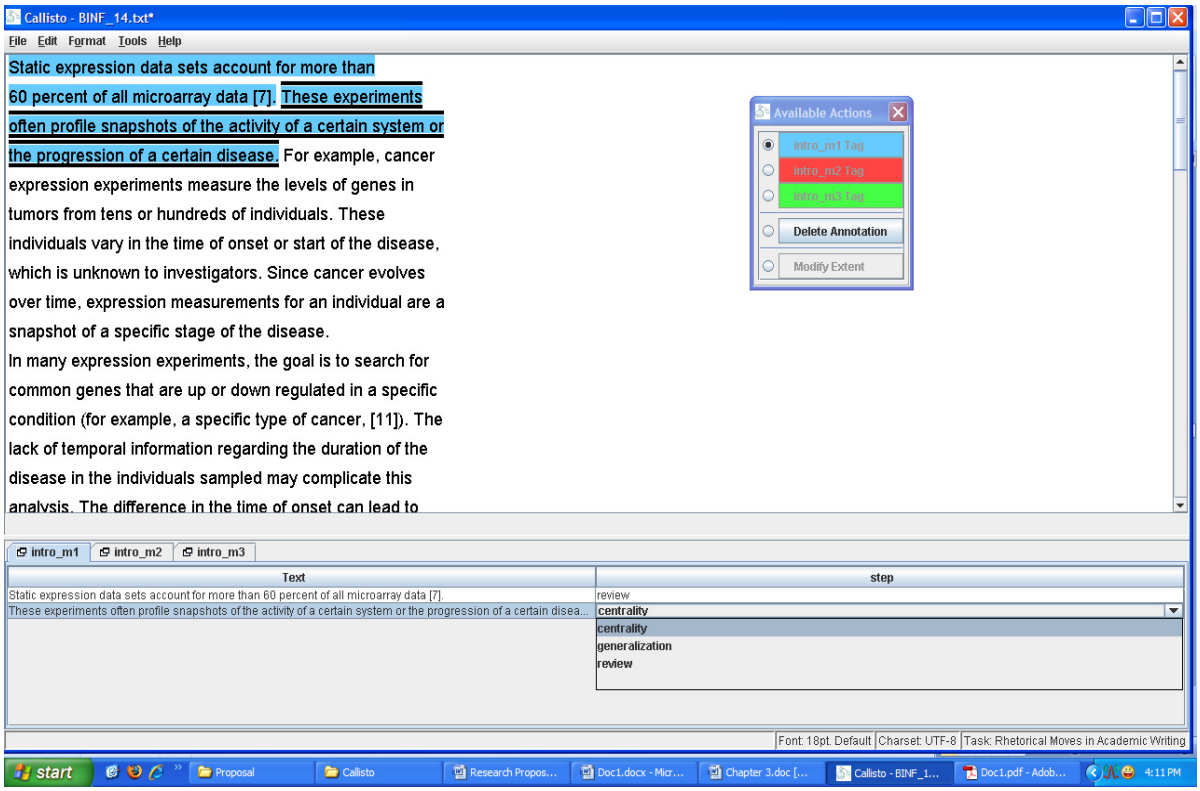


Figure 3.15. Example of text annotation in Callisto

For this corpus annotation, inter-annotator agreement was calculated. Two annotators annotated a random sample of research articles introductions, which amounted to 487 sentences. The inter-annotator agreement Kappa (*k*), which is the probability of agreement minus chance agreement, was calculated as follows:

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}$$

where P(A) represents observed probability of agreement, and P(E) is the expected probability of agreement, i.e., chance agreement. Given three moves and uniform distribution among them, $P(E) = (\frac{1}{3})^2$. The two annotators had an average *k* of 0.945 over the three

moves. Table 3.2 shows the relatively high agreements between the two annotators for each move. Agreement between human raters and IADE is reported in Section 4.6.5.2.

Table 3.2. Inter-annotator agreement

	Move 1	*Move 2*	*Move 3*
N. agreed	457	452	480
P(A)	0.938	0.928	0.986
k	0.931	0.919	0.984

N. agreed – number of sentences on which both annotators agreed (out of 487 sentences)
P(A) – observed probability of agreement
k – inter-annotator agreement

The annotated corpus was needed to build IADE's analysis module that performs automatic identification of discourse moves in RA Introductions. This task was approached as a classification problem. In other words, given a sentence and a finite set of moves and steps, what move/step does the sentence signify? This task is very similar to identifying the discourse structure of short argumentative essays discussed in Burstein et al. (2003), the difference being in the genre of the essays and type of the discourse functions in question.

- Feature selection

In order to classify sentences correctly, first features that can reliably indicate a move/step needed to be identified. A text-categorization approach was taken to address this problem (see Pendar & Cotos, 2008). In this framework, each sentence is treated as a data item to be classified, and is represented as an n-dimensional vector in the $\mathcal{R}^n$ Euclidean space. More formally, a sentence *Si* is represented as the vector

$\bar{s}_i = \langle f_1, f_2, \ldots, f_n \rangle$ where each component f_j of the vector $\bar{s}_i$ represents a

measure of feature j in the sentence s_i. The task of the learning algorithm is to find a function $F : S \rightarrow C$ that would map the sentences in the corpus (S) to classes (C) in M = {m1,m2,m3} (where m1, m2, and m3 stand for Move 1, Move 2, and Move 3, respectively).

An important problem was choosing features that would allow for proper classification of the data instances into the classes in question. For this purpose, word unigrams, bigrams, and trigrams (i.e., single words, two word sequences, and three word sequences) were extracted from the annotated corpus. It is worth mentioning that not all the n-grams were used. The low frequency n-grams at a cut-off point of 5 were removed from the data pool. The total number of each remaining set of *n*-grams used is shown in Table 3.3.

Table 3.3. Total number of extracted *n*-grams

n-gram	*Number*
unigrams	3,951
bigrams	8,916
trigrams	3,605

To identify which *n*-grams are better indicators of moves, odds ratios were calculated for each as follows:

$$OR(t_i, m_j) = \frac{p(t_i|m_j) \cdot (1 - p(t_i|\bar{m}_j))}{(1 - p(t_i|m_j)) \cdot p(t_i|\bar{m}_j)}$$

where $OR(t_i, m_j)$ is the odds ratio of the term (*n*-gram) t_i occurring in move m_j; $p(t_i|m_j)$ is the probability of seeing the term t_i given the move m_j; and $p(t_i|\bar{m}_j)$ is the probability of seeing the term t_i given any move other than m_j. The above conditional probabilities were calculated as maximum likelihood estimates.

$$p(t_i|m_j) = \frac{\text{count}(t_i \text{ in } m_j)}{\sum_{k=1}^{N} \text{count}(t_k \text{ in } m_j)}$$

where N is the total number of *n*-grams in the corpus of sentences S. The terms with maximum odds ratios were selected as features. These features were then used to build IADE's classifier (see Pendar & Cotos (2008) for details).

The discourse analysis and annotation was a prerequisite not only for training and developing the analysis module of the program, but also the source of the color-coded feedback generated by it (Figure 3.14a). The content of the numerical feedback (Figure 3.14b) draws from statistical analyses on moves' length and frequency in the annotated corpus for each discipline. Similar statistics were calculated for the purpose of providing the students with a more detailed description of the discourse development of RA introductions at the level of steps in IADE's Help Options (see Figure 3.12).

3.4. Chapter summary

Given that the major goal of this study was to evaluate the effectiveness and appropriateness of IADE for formative assessment purposes, and because context is crucial

in formative assessment, this chapter overviewed the academic writing instructional context in which IADE was implemented. This context is the English 101D course for international graduate students, for whom it is a requirement that results from their performance on the institutional English Placement Test. The English 101D pedagogy employs a genre and corpus-based approach, focusing on the research article genre and using a corpus of research articles published in various disciplines. This course makes use of a number of computer applications such as Moodle, a course management system, and WordSearch, a concordancer. In the teaching of research article introductions, which is of interest in this study, additional computer-based materials are offered. Students conduct analyses of an annotated corpus of introductions in their discipline with the help of the Calisto software, which displays the rhetorical moves through colors and the steps within the moves through notations for each sentence. Then they write a draft of an introduction for their own research article, and complete its revision with feedback from IADE. Because IADE is a new AWE program developed specifically for the English 101D context, this chapter explained how its design is grounded in practice, theory, technology, and research. The next chapter provides a detailed account of the research methodology employed in the empirical evaluation of IADE.

Chapter 4. Methodology

This chapter delineates the methodology undertaken in this dissertation, which follows a process-product approach to evaluate the use of technology as a supplement to instruction. It employs the mixed-methods concurrent transformative strategy by integrating both quantitative and qualitative data collected from a total of 105 participants. The characteristics of the participants, who are international graduate students enrolled in English 101D, are described based in part on the information provided by the students themselves. Following that, a detailed account of the materials and data collection instruments is provided. IADE is described from two perspectives - that of data collection and that of implementation.

In addition, the chapter recounts the particulars of other data sources such as pre- and post-tests; Likert-scale, yes/no, and opened-ended survey questions; think-aloud protocols; observations; and interviews. The Camtasia and Audacity technologies that were used to capture the data are also presented and exemplified here, as is Transana that was used to code and analyze most of the qualitative data. Further, essential is the description of the research procedure, which was the same in terms of IADE implementation, but somewhat differed in terms of data collection over two semesters during which the study was conducted. Finally, following the description of the procedure, the data sources and data analysis are elaborated. The explanation of data analysis is developed according to the four CALL qualities investigated: Language Learning Potential, Meaning Focus, Learner Fit, and Impact, for which both quantitative and qualitative analyses are delineated.

4.1. Approach

This study follows a call for the initiation of a classroom research agenda that would help evaluate and guide the application of automated evaluation in the writing pedagogy, which has been made by Warschauer and Ware (2006). Based on a categorization developed by Long (1984), they propose three directions for research: product, process, and process/product, where "*product* refers to educational outcome (i.e. what results from using the software), *process* refers to learning and teaching process (i.e. how the software is used), and *process/product* refers to the interaction between use and outcome" (p.10). This study pursued the process-product research direction (Warschauer & Ware, 2006)

On the level of evaluating technology for language learning in general, Chapelle (2007) specifies three targets for evaluation: "what is taught in a complete course", "what is taught through technology in a complete course", and "what is taught through technology" (p. 30). In the first case, an entire technology-based course is evaluated, in the second case – CALL materials used for learning a subset of course objectives, and in the third case – the use of technology as support and enhancement of a face-to-face course. The main implication for recognizing these types of CALL uses is predicating the difference in the need for different evaluation approaches. In this study, the second target for evaluation specified by Chapelle (2007) was chosen as the focus of evaluation since IADE represents CALL materials used for learning of rhetorical moves in research article introductions, which are one of the English 101D course objectives. It is a very complex type of approach because the technology cannot be isolated to investigate the effects.

Given a great degree of complexity and given that this program represents a case of innovative technology, its evaluation, as well as the evaluation of any other new CALL applications, according to Chapelle (2007), is "perhaps the most significant challenge teachers and curriculum developers face when attempting to introduce innovation into language education" (p. 30). The innovation here is AWE in the form of IADE, and its implementation requires multi-faceted empirical evidence about its effectiveness. To obtain such evidence, this study employed a mixed methods approach.

The mixed methods approach originated in psychology research (Campbell & Fiske, 1959), but has been gaining more and more recognition in Applied Linguistics. It is defined as "one in which the researcher tends to base knowledge claims on pragmatic grounds" and in which data collection "involves gathering both numeric information […] as well as text information […] so that the final database represents both quantitative and qualitative information" (Creswell, 2003, pp. 19-20). This study followed the concurrent transformative strategy that is typically guided by theory, which in this case is instructed SLA represented by Chapelle's (2001) CALL evaluation conceptual framework. Figure 4.1 shows how the framework overarches the chosen methodological approach. Both quantitative and qualitative data were collected and integrated during the analysis and interpretation phases. The quantitative data consisted of Likert-scale and yes/no survey responses, scores that measured learning gains based on pre-/post-tests, scores that measured discourse improvement based on automated and human ratings, and frequencies of submission and access to IADE's Help Options. The qualitative data consisted of open-ended survey responses, participants' first and last drafts, screen capturing as well as recordings of think-aloud protocols, observations,

and interviews. These data were collected during the course of two semesters, the data collection procedures being somewhat different, which is further described in Section 4.4. Priority was given to quantitative evidence for purposes of possible generalization of findings. The reason for employing a mixed-methods design that converges findings from various sources of evidence was the intent to fully understand the nature of factors that can potentially shed light on the formative effectiveness of IADE and its feedback evaluated in terms of its CALL qualities.

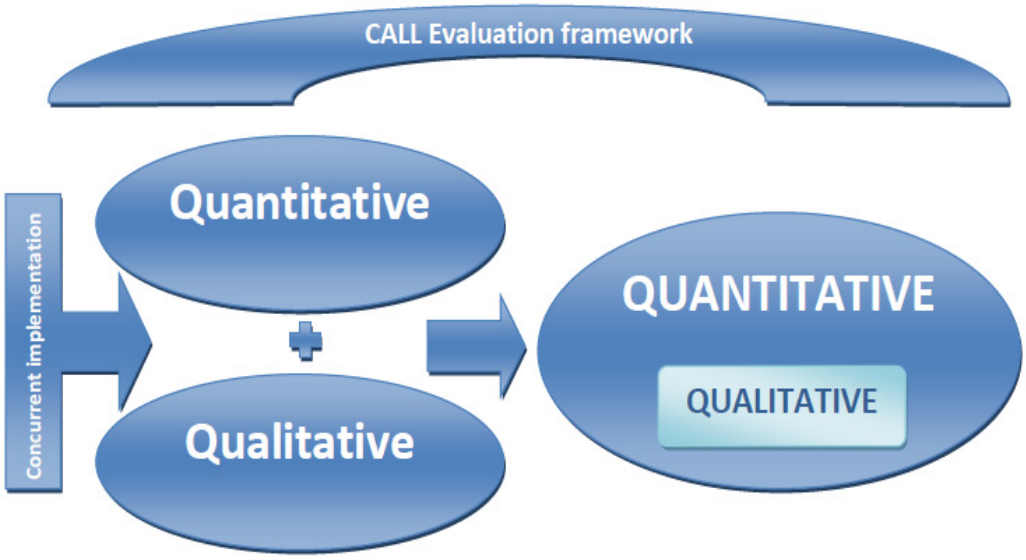


Figure 4.1. Concurrent transformative strategy (based on Creswell, 2003)

4.2. Participants

The participants in this study were 105 international students (59 male and 46 female) at Iowa State University who were registered in six sections of the English 101D writing course in the Fall of 2008 and the Spring of 2009. They were all enrolled in graduate programs at this university. Thirty seven of them were Masters students, and sixty six were PhD students specializing in one of the following disciplines:

1.Accounting
2.Aerospace Engineering
3.Agronomy
4.Analytical Chemistry
5.Animal Science
6.Biochemistry
7.Bioinformatics
8.Biology
9.Biomedical Science
10. Bionanotechnology
11. Chemical Engineering
12. Computer Engineering
13. Computer Science
14. Curriculum and Instruction
15. Economics
16. Electrical Engineering and Power Systems
17. Environmental Engineering
18. Ergonomics
19. Food and Lodging Management
20. Genetics
21. Human Health and Public Performance
22. Industrial Engineering
23. Journalism
24. Materials Science and Engineering
25. Mathematics
26. Mechanical Engineering
27. Physics and Astronomy
28. Plant Breeding
29. Public Administration
30. Sociology
31. Statistics
32. Textiles and Clothing
33. Urban and Regional Planning
34. Veterinary Medicine

Table 4.1 summarizes the characteristics of participating students based on information obtained from three data sources (IADE's database, the survey, and the Likert-scale questionnaire) introduced in the following section. The participants were aged between 22 and 44. They represented various language backgrounds, most of the students' first language being Chinese, and the rest speaking either Thai, Italian, Spanish, Korean,

Portuguese, Arabic, Turkish, Telugu, Filipino, or Greek. English was a second language for 77 participants. 24 participants spoke two languages other than their native and English, and 2 spoke three other languages. The participants' overall English language proficiency is reflected in their TOEFL scores, which ranged between 80 and 104 on the Internet-based test, between 243 and 255 on the computer-based test, and between 520 and 667 on the paper-based test. Eighty eight participants self-evaluated the level of their writing skills by choosing one option in the first Likert-scale question:

What is your level of English writing skills?

Excellent very good good fair poor

None of the students evaluated their writing as 'excellent'. Five students evaluated themselves as 'very good'; 33 as 'good'; 41 as 'fair', and 9 as 'poor'.

Their period of study at a US university ranged from a few months to two years. For 65 students, it was the first semester at a US university, for 19 – the second, for 10 – the third, for 4 – the fourth, and for 5 – the fifth. Given that English 101D targets the research article genre, of interest was their experience of writing research articles. Most of the students (84) had not written a research article before signing up for this course. Forty three students had written research articles in their native language, only twenty two of whom had published their papers. Nineteen students had written research articles in English, although 14 of them had produced them as part of their course work without the intent to pursue publishing. Five students had research articles published in English; however, those articles were results of their collaborative work with faculty, which is typical in most fields. Only one student was the sole author of his publication.

Table 4.1. Participants' characteristics

Age (students)	*L1* (students)	*Other lang-s* (students)	*Semesters at US univ.* (students)	*TOEFL scores (test type)*	*RA in L1*		*RA in English*	
					Yes	*No*	*Yes*	*No*
22 (9) 23 (8) 24 (10) 25 (15) 26 (14) 27 (6) 28 (7) 29 (5) 30 (4) 31 (9) 32 (2) 33 (4) 34 (3) 35 (2) 37 (3) 38 (1) 39 (1) 40 (1) 44 (1)	Chinese (72) Thai (1) Italian (1) Spanish (7) Korean (11) Portuguese (2) Arabic (3) Turkish (4) Telugu (2) Filipino (1) Greek (1)	1 (79) 2 (24) 3 (2)	1 (67) 2 (19) 3 (10) 4 (4) 5 (5)	80-104 (iBT) 243-255 (CBT) 520-667 (PBT)	43 (22 no/pub)	62	19 (14 no/pub)	86

4.3. Materials and instruments

As the evaluation of IADE in the context of an academic writing course was the purpose of this study, it was the core material used. IADE was used both as a complementary formative assessment tool in instruction and as a data collection instrument whose database recorded automated analysis data as well as student drafts and frequencies of submission and access to Help Options. Data were also collected by means of a number of other instruments. More quantitative data were gathered through pre- and post-tests and Likert-scale, yes/no, and open-ended survey questions (responses to the latter were used as a source of qualitative data). Other sources of qualitative data were think-aloud protocols, semi-structured interviews, and observations, which were elicited from the same sample of 16 participants.

4.3.1. IADE's database

IADE, the AWE application described in Chapter 3, has a database (or the Evidence Composite Library component, as referred to in Section 3.1.3) with the help of which the following types of information were captured:

1. Students' first and last names
2. Students' automatically assigned ID numbers[1]
3. Students' classification (PhD or MA)
4. Students' disciplines
5. All the drafts submitted by each student
6. The automated analysis and feedback generated by the program
7. The number of hits by individual students on each Help Option in the program
8. The date and time of draft submission and of access to Help Options.

This information was accessed as an online webpage, where it was displayed in three parts. The first was the so-called 'Corpus tracking', which includes automatically assigned student IDs, students' first and last names, students' classification into masters or PhD, ID numbers for students' drafts, names of annotated corpus documents, and the number of hits on those. A screenshot of the original online database taken during testing by non-participants is shown in Figure 4.2a.

[1] These ID numbers were used during the analysis phase and will be accompany the quotes provided to exemplify the findings presented in the results Chapters 5 and 6.

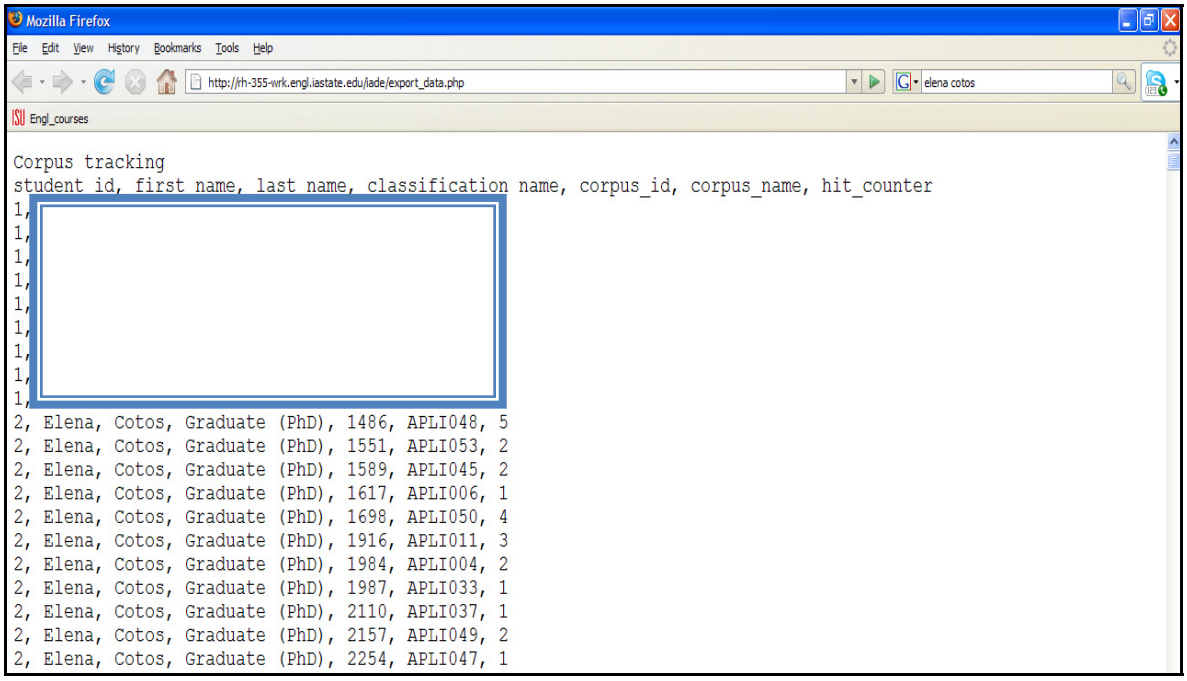


Figure 4.2a. The first part of the data displayed as a webpage

The second part, under a 'Help tracker' heading, in addition to automatically assigned student IDs, students' first and last names, students' classification into masters or PhD, and the number of hits, included information about the Help Options accessed in IADE and the time of access. For instance, as exemplified in Figure 4.2b, the database shows that the second user (Elena Cotos) clicked on different Help Options at different times: 15 times on the definitions of moves/steps on 2009-04-02 at 10:26:49, and around the same time she clicked 12 times on step statistics, 4 times on revision tips, and 15 times on the annotated corpus. The lines below that information show that the database also captured the same kind of data for this users' access to individual definitions of moves and steps.

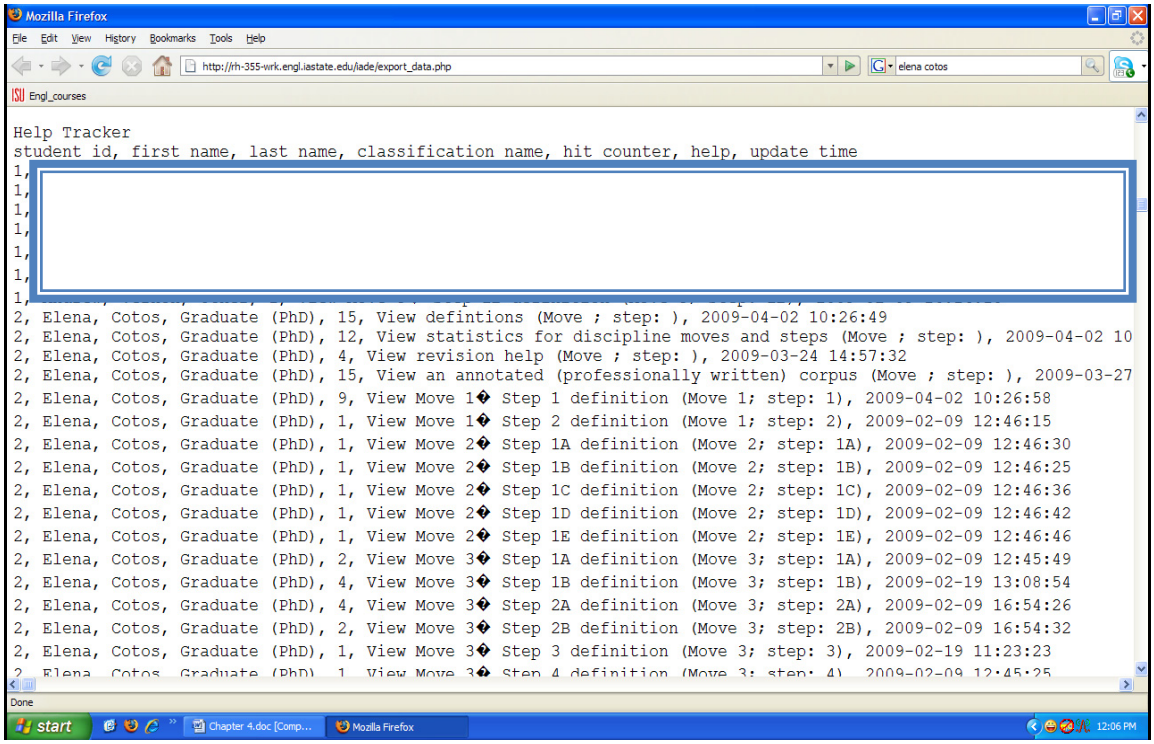


Figure 4.2b. The second part of the data displayed as a webpage

Similarly, the third part, called 'Student written corpora', captured not only students' personal information, but also all students' drafts, the number of their submission, and, most importantly, the automated analysis of each draft. As it can be seen in Figures 4.2a, 4.2b, and 4.2c, all the information for each student entry in IADE was displayed as one line, which in most of the cases made it impossible to clearly see and work with the data. Consequently, the data accessed in this online form was copy-pasted into Notepad and then saved as a .csv file. The .csv file format allowed for saving the entire data set into one Excel document, where every information category (i.e., student ID, first name, last name, degree, discipline, number of drafts, time) occupied a separate cell and could be seen in its full form by double-clicking on a selected cell. A sample screenshot is given in Figure 4.3, where the names of the participants are covered by white boxes.

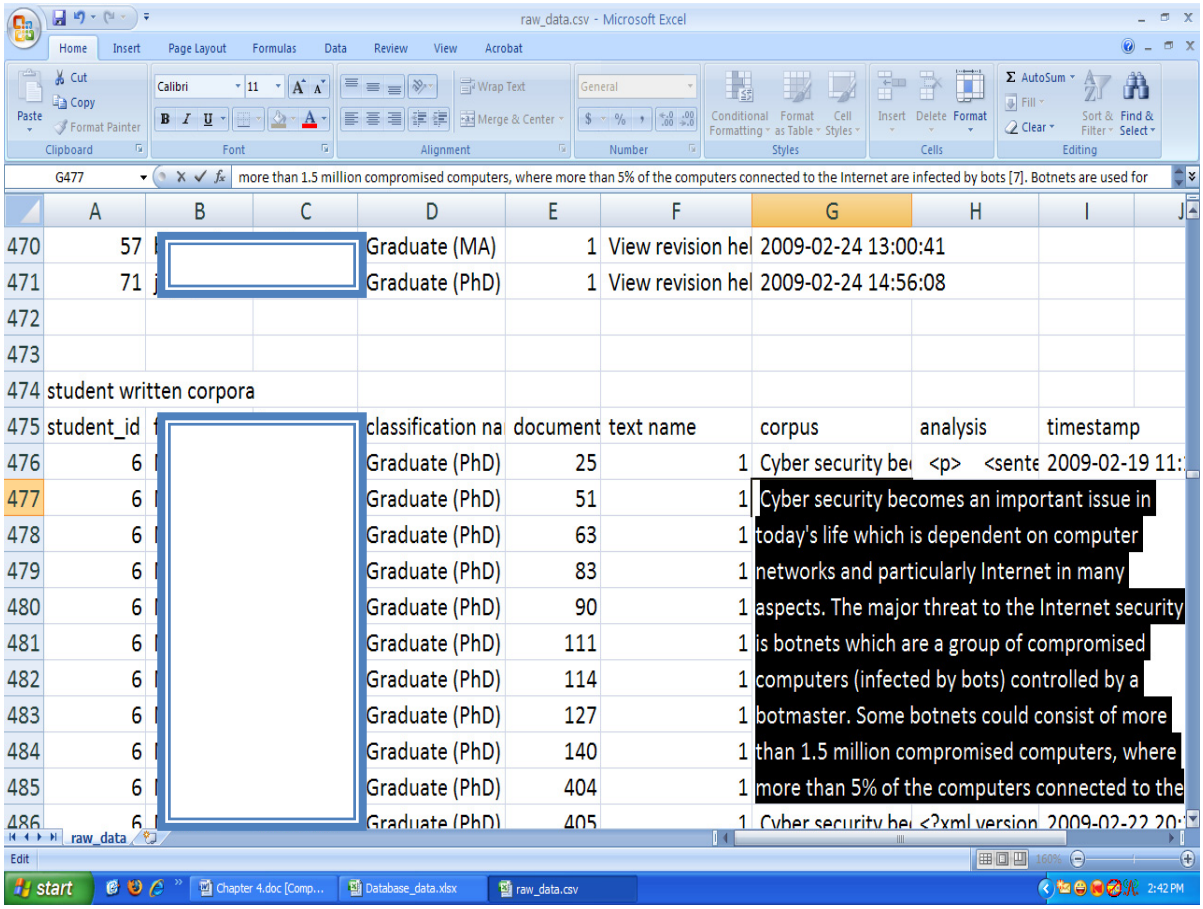
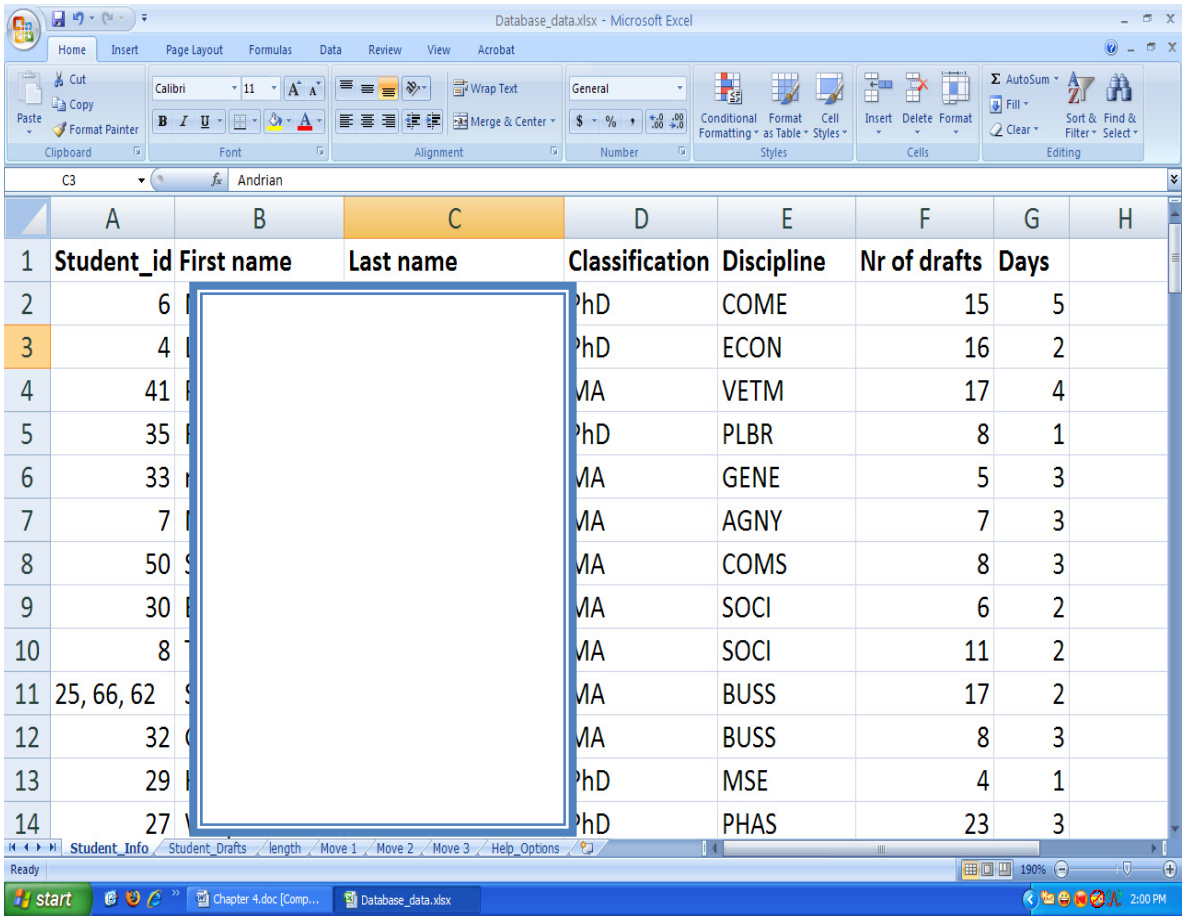


Figure 4.3. Visualization of IADE's database saved in Excel

Because all the three parts of the database were imported as one big worksheet, separate worksheets were then created by the researcher for particular types of data to facilitate their analysis. Specifically, worksheets were created with the following data: student information, student drafts, drafts' length, Move 1, Move 2, Move 3, and Help Options. The data worksheet with the student information, exemplified in Figure 4.4a, contained the student IDs, students' first and last names, students' classification into masters or PhD, their discipline, the number of submitted drafts, and the number of days each student worked with IADE.

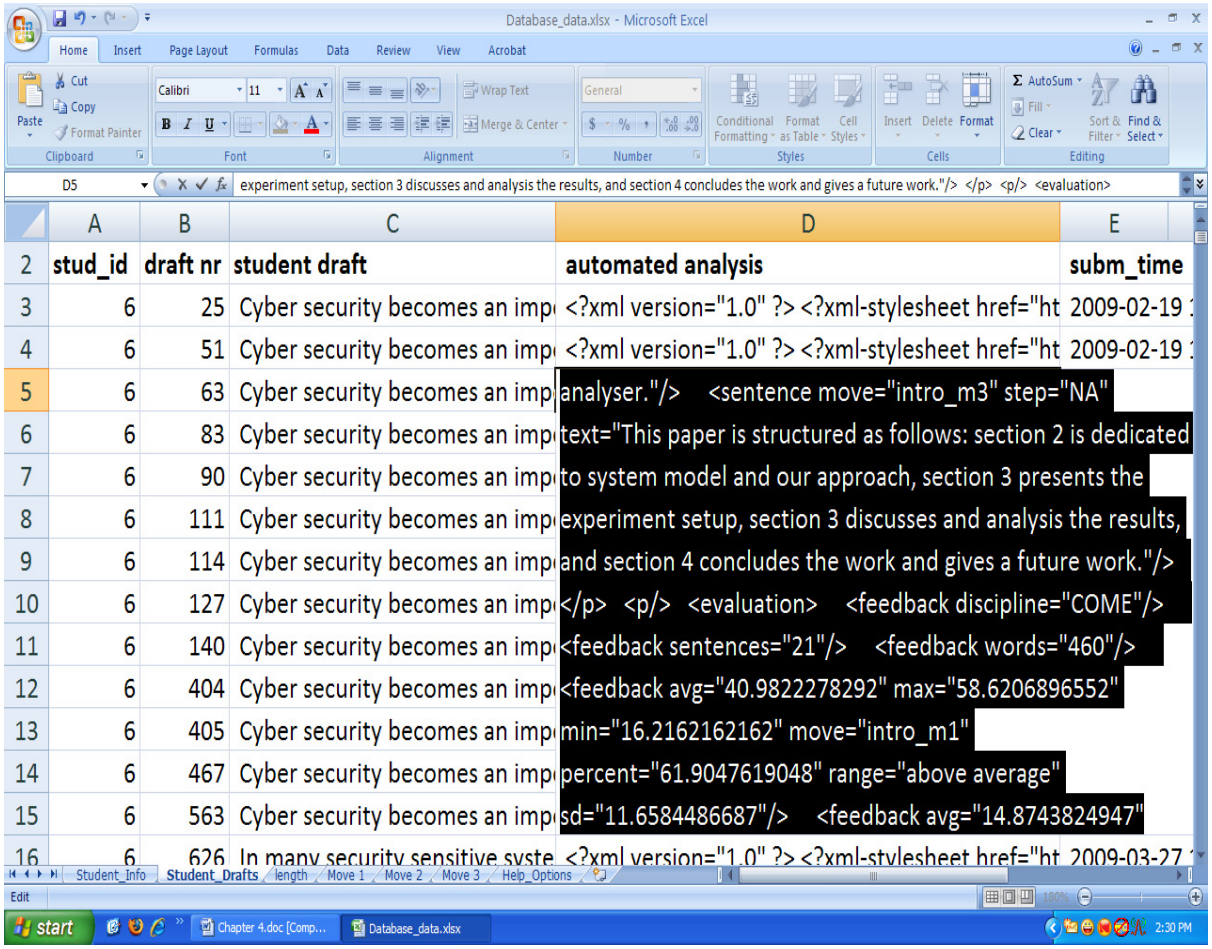


	A	B	C	D	E	F	G
1	Student_id	First name	Last name	Classification	Discipline	Nr of drafts	Days
2	6			hD	COME	15	5
3	4			hD	ECON	16	2
4	41			MA	VETM	17	4
5	35			hD	PLBR	8	1
6	33			MA	GENE	5	3
7	7			MA	AGNY	7	3
8	50			MA	COMS	8	3
9	30			MA	SOCI	6	2
10	8			MA	SOCI	11	2
11	25, 66, 62			MA	BUSS	17	2
12	32			MA	BUSS	8	3
13	29			hD	MSE	4	1
14	27			hD	PHAS	23	3

Figure 4.4a. Student information worksheet in the Excel database

Note: Line 11 has three ID numbers for the same participant. This participant forgot the log-in password and opened new accounts when re-entering IADE.

The worksheet with the student drafts had three main data columns: student ID, student draft, automated analysis, and time of draft submission. The full form of a student draft as well as of the analysis of that draft captured in IADE's database were exported into a single cell, like all the other stored information, and could be viewed by double-clicking on the respective cell (Figure 4.4b).



	A	B	C	D	E
2	stud_id	draft nr	student draft	automated analysis	subm_time
3	6	25	Cyber security becomes an imp	<?xml version="1.0" ?> <?xml-stylesheet href="ht	2009-02-19
4	6	51	Cyber security becomes an imp	<?xml version="1.0" ?> <?xml-stylesheet href="ht	2009-02-19
5	6	63	Cyber security becomes an imp	analyser."/> <sentence move="intro_m3" step="NA"	
6	6	83	Cyber security becomes an imp	text="This paper is structured as follows: section 2 is dedicated	
7	6	90	Cyber security becomes an imp	to system model and our approach, section 3 presents the	
8	6	111	Cyber security becomes an imp	experiment setup, section 3 discusses and analysis the results,	
9	6	114	Cyber security becomes an imp	and section 4 concludes the work and gives a future work."/>	
10	6	127	Cyber security becomes an imp	<evaluation> <feedback discipline="COME"/>	
11	6	140	Cyber security becomes an imp	<feedback sentences="21"/> <feedback words="460"/>	
12	6	404	Cyber security becomes an imp	<feedback avg="40.9822278292" max="58.6206896552"	
13	6	405	Cyber security becomes an imp	min="16.2162162162" move="intro_m1"	
14	6	467	Cyber security becomes an imp	percent="61.9047619048" range="above average"	
15	6	563	Cyber security becomes an imp	sd="11.6584486687"/> <feedback avg="14.8743824947"	
16	6	626	In many security sensitive syste	<?xml version="1.0" ?> <?xml-stylesheet href="ht	2009-03-27

Figure 4.4b. Student drafts worksheet in the Excel database

The data worksheet on the length of student drafts, also containing students' IDs, included the counts for the number of words in the first and in the last drafts, the word average in each student's discipline, as well as the comments (e.g., 'about average', 'far above average', 'below average', etc.) that were automatically generated by the program based on the comparison of the length of a student draft with the average length of research article introductions in his/her discipline (See Figure 4.4c). All these numbers and the comments were extracted from the student drafts worksheet presented above.

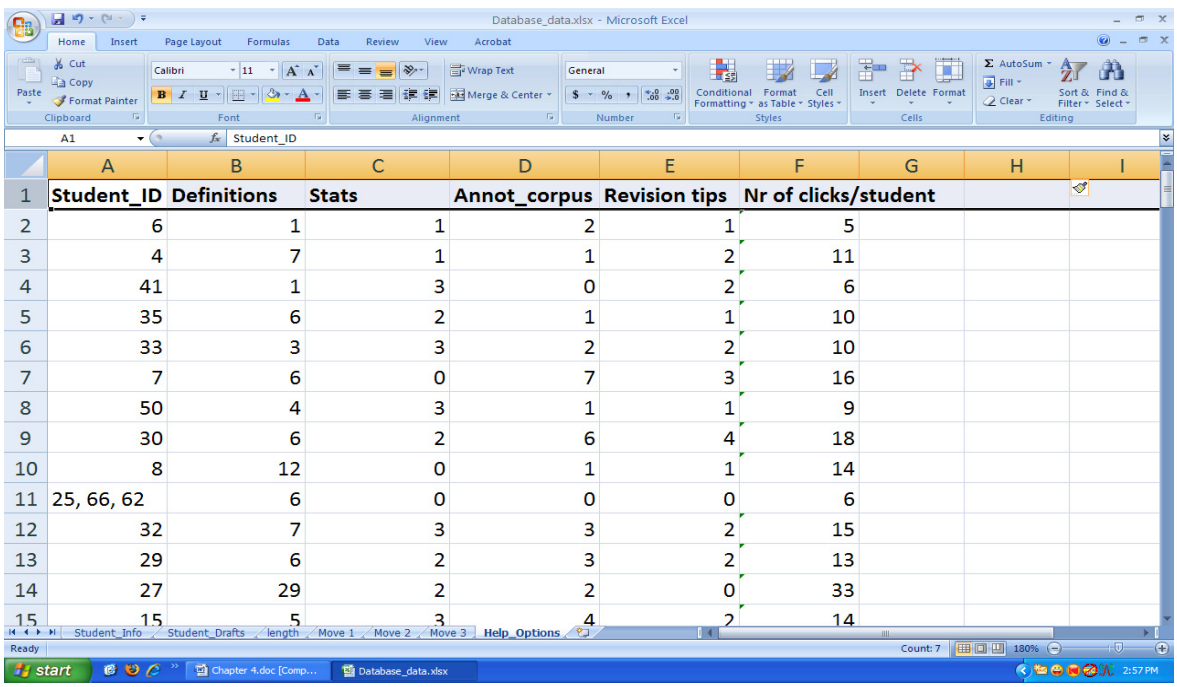
	A	B	C	D	E	F
1	Student_id	First draft words	Last draft words	Words average	First draft comment	Last draft comment
2	6	479	404	625	about average	about average
3	4	1937	1936	717	far above average	far above average
4	41	371	430	656	below average	below average
5	35	501	526	600	about average	about average
6	33	195	395	804	far below average	below average
7	7	941	971	605	above average	far above average
8	50	298	540	1058	below average	about average
9	30	529	540	1058	about average	about average
10	8	540	464	631	about average	about average
11	25, 66, 62	401	420	631	about average	about average
12	32	339	487	631	below average	about average
13	29	533	551	435	about average	about average
14	27	205	272	687	below average	below average

Figure 4.4c. Data worksheet on draft length

	A	B	C	D	E
1	Student_id	First draft Move 1	Last draft Move 1		
2	6	far above average	about average		
3	4	about average	about average		
4	41	far below average	about average		
5	35	below average	about average		
6	33	about average	about average		
7	7	above average	about average		
8	50	about average	about average		
9	30	above average	about average		
10	8	about average	about average		
11	25, 66, 62	above average	about average		
12	32	about average	about average		
13	29	above average	about average		
14	27	below average	about average		
15	15	about average	about average		

Figure 4.4d. Data worksheet on Move 1

Similar data extraction was done to create the worksheets for the three moves, where, next to each student ID, the evaluation of each Move in the form of IADE analyzer's comments was introduced for the first and the last drafts (see an example in figure 4.4d). The reason for singling out the comments is that they were needed for the further purpose of assigning scores to the students' texts. The last worksheet consisted of numerical information only. Specifically, the frequency with which the students accessed each of the four Help Options available in IADE was recorded here.



Student_ID	Definitions	Stats	Annot_corpus	Revision tips	Nr of clicks/student
6	1	1	2	1	5
4	7	1	1	2	11
41	1	3	0	2	6
35	6	2	1	1	10
33	3	3	2	2	10
7	6	0	7	3	16
50	4	3	1	1	9
30	6	2	6	4	18
8	12	0	1	1	14
25, 66, 62	6	0	0	0	6
32	7	3	3	2	15
29	6	2	3	2	13
27	29	2	2	0	33
15	5	3	4	2	14

Figure 4.4e. Data worksheet on Help Options

4.3.2. Pre- and post-tests

To measure the students' knowledge of the moves and of the steps before and after their interaction with IADE, a pre-test and a post-test were designed. Both tests consisted of two tasks (see Appendix A). The first task required the students to name the moves and the steps in a number of given decontextualized examples. The second task focused on

annotating each sentence of a research article introduction in terms of moves and steps by using the 'Insert comment' function in Word documents. Here, being given the entire text, the students could make their judgments about the function of each sentence based on the context. The number of sentences in both tasks of the pre- and post-tests was the same – 17. The pre- and post-tests differed because they contained different sentences and different texts; however, their content was parallel. In other words, the sentences in task one were representative of all the moves and their steps, and the texts in task two contained all the three moves and almost the same steps.

It should be mentioned here that finding two authentic texts that would consists of exactly the same steps distributed in the exactly the same way proved to be very difficult, which is why the introductions selected for the pre- and post-tests, although very similar, differed in the lack of one step of Move 3 (Outlining the structure of the paper) in the pre-test text. In addition to the presence of the same moves and steps, another important criterion for text selection was comprehensibility, that is, the meaning of the texts had to be clear and not burdened by discipline-specific terminology. Both texts that were included in the tests were in Economics and proved to satisfy this requirement after being piloted with a group of 17 students who had taken English 101D one semester prior to this study.

4.3.3. Yes/No and open-ended survey questions

The survey (see Appendix B) contained 8 questions eliciting such information about the students' characteristics as students' age, gender, first language, knowledge of other languages, period of study at a US university prior to taking English 101D, TOEFL score, and research article writing experience. Twenty five yes/no questions were geared towards

the four CALL qualities under investigation. The first twelve elicited information related to the Language Learning Potential of IADE by inquiring about students' focus on moves or other relevant discourse elements as well as about their perceptions of learning and self-improvement. The Learner Fit quality was addressed by five questions, through which the students were asked to evaluate the appropriateness and helpfulness of different features of IADE. Three other questions meant to find evidence of Meaning Focus, or whether and how the students had a focus on the meaning of their discourse while revising with IADE. Evidence of the fourth quality, Impact, was elicited through five questions that sought data about students' positive and negative experiences during their interaction with the program. All the questions in the survey were accompanied by open-ended follow-up questions such as "Why?", "How?", or "What makes you say that?" Participants' responses to the open-ended questions were analyzed by identifying emerging themes, which were then quantified in terms of percentages of students who mentioned them.

4.3.4. Likert-scale survey questions

The Likert-scale questions complemented other data sources with 19 questions (Appendix C). The first question asked the students to self-assess their general level of English language proficiency on a scale of "excellent", "very good", "good", "fair", and "poor". The rest of the questions offered a choice of four answers, which, depending on the question, were: "a lot" or "very well", "somewhat" or "well", "a little", and "not at all". A certain number of questions required the evaluation of particular aspects of the four CALL qualities analyzed: seven questions for Language Learning Potential, three questions for Meaning Focus, five questions for Learner Fit, and four questions for Impact. One question

for Learner Fit, however, comprised ten items related to specific characteristics of IADE's feedback (explicit, individual specific, metalinguistic, color-coded, negative, intelligent, output-focused, iterative, and short). Participants' choices on individual scale questions were converted to percentages. Also, for the ultimate purpose of overall evaluation of a given CALL quality, the four choices were equaled to a certain degree of evidence strength. In other words, "a lot" or "very well" was considered as excellent evidence of a CALL quality, "somewhat" or "well" as good evidence, "a little" as weak evidence, and "not at all" as poor evidence.

4.3.5. Think-aloud protocols

During the first semester of data collection, 16 out of 53 participating students volunteered to interact with IADE in a controlled experimental setting in a computer lab outside of class. Only the participating student and the researcher were in the lab. For the purpose of introspective data collection, they were asked think aloud (Ericsson & Simon, 1987), i.e., to say whatever they were thinking about, whatever they were looking at, whatever they were feeling, and whatever they were doing to go about the task. The think-aloud time ranged from 25 to 37 minutes. Before the participants began revising with the program, to ensure that they understood this unusual task, they were given a written explanation and instructions to read (Appendix D). Also, for the same purpose, a short demonstration of a think aloud protocol was performed by the researcher for each student. During the think-aloud protocols, the researcher also prompted the students to verbalize their thoughts whenever it was necessary, ensuring that they produce data similar to the example in Figure 4.5. Here, Student 27 noticed that a sentence, in which he referred to previous

research and which was meant as Move 1, was identified by IADE as Move 3. Having noticed that, he verbalized his thoughts, asking himself why this happened and speculating about possible reasons, which in this case he thought was the placement of the sentence in the text. Then, the student stated aloud what he intended to do next, i.e., to try to change the position of that sentence in order to observe whether that would make a difference.

> There is a question that for m1, my previous research was identified as m3. Why it happened like this? I think maybe it's the position, I guess, maybe the program detects according to its position… if it appears at the beginning of this paragraph, so I guess maybe the program just sees it as m1 according to its location. Let's see if I change the position, what will happen. (Student 27, think-aloud protocol)

Figure 4.5. Excerpt from a think aloud protocol

The audio recording function of Camtasia Studio 5 software by TechSmith was used to unobtrusively record participants' think-aloud protocols. A microphone was plugged into the computer on which the participant was working. Camtasia's screen recording function was also used to capture the data on participants' interaction with IADE in the form of video files, which contained the recording of all the actions visible on the participant's computer screen (see Figure 4.6). The video files created in Camtasia were about 250 - 400 MB each and were saved in .avi format because this format was necessary to import the think-aloud and screen recording data into Transana.

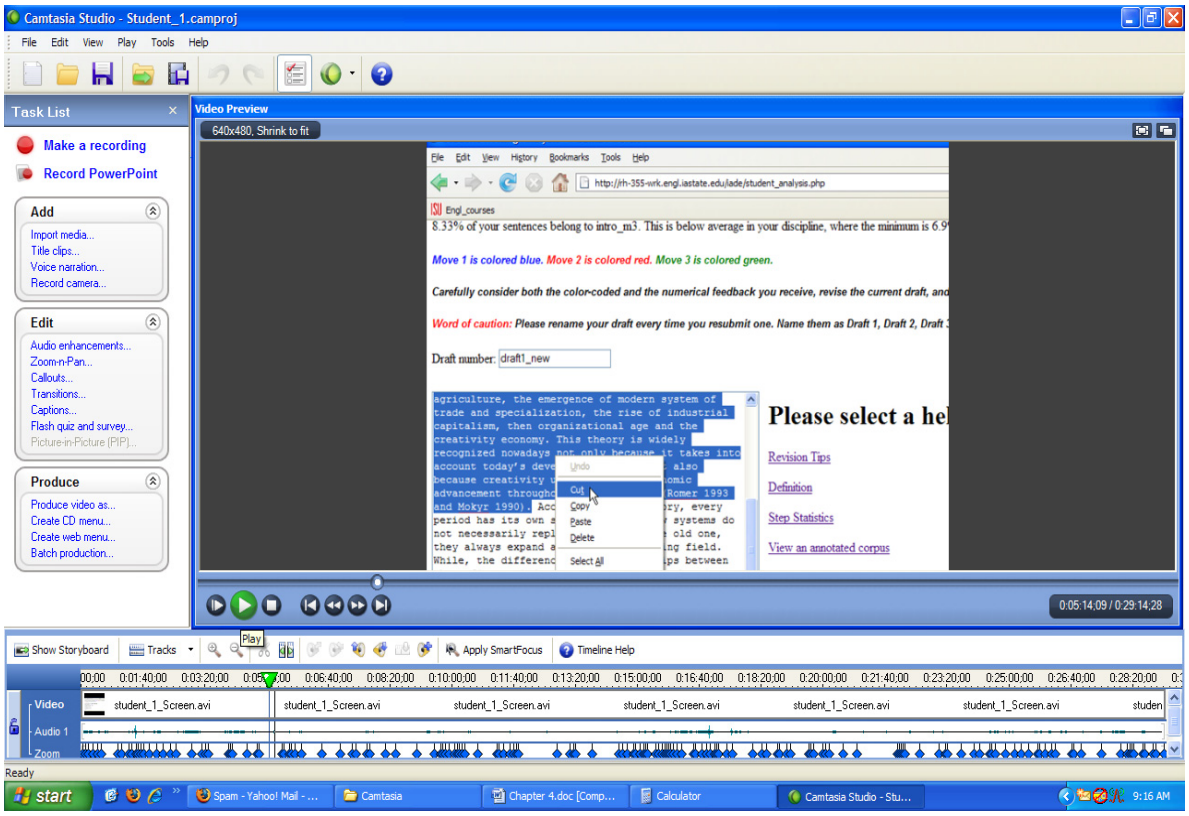


Figure 4.6. Screenshot from a Camtasia file

4.3.6. Observations

The researcher conducted observations of the same 16 students who participated in the think-aloud protocol data collection. The observations were conducted during each revision session in the same lab conditions. The researcher sat behind the participant at a distance from which she could see the student's computer screen. Notes about each participant's behavior during the interaction with IADE were made on paper, and question marks were put next to the entries that required further clarification, which was to be made during the interview shortly after the revision session was over. The observation notes were then typed on the computer and saved as rich text format documents so that the transcripts could be imported into Transana. The length of the observation notes ranged from one to two

pages of 12 Times New Roman single-spaced text. Figure 4.7 shows an excerpt from the observation notes on Student 28. The meaning of the question marks is explained in Section 4.3.7 below.

> The first time he receives the feedback, he looks at one color then goes back to the explanation materials to check (ok, this is move 1, and move 1 is claiming centrality). He does the same for the other two moves as well.
>
> He notices that there is a discrepancy between what he perceives as a particular move and what the program outputs. **(?)**
>
> Then he thinks more about what he wrote and how it sounds (he keeps reading and re-reading those pieces of text), and decides that there is indeed a reason for why the program shows what it shows. **(?)**
>
> He decides to revise. **(?)**
>
> Even though the numerical feedback showed that all his moves were about average compared to the introductions in his field, he still decided to read carefully and make changes. **(?)**
>
> He frequently turns to Help Options, but mostly with the purpose of seeing examples. He changes one part of speech for another.
>
> Reads aloud a lot and thinks about how the meaning he is trying to convey would be understood by the program.
>
> He appears to be deeply cognitively engaged. **(?)** (Student 28, observation notes)

Figure 4.7. Excerpt from observation notes

4.3.7. Interviews

Immediately after the revision sessions, 10-15 minute retrospective semi-structured interviews were conducted with all the 16 participants in the same computer lab. The interviews are considered semi-structured because the questions asked were about those

actions and/or utterances that were marked with a question mark in the observation notes (see examples in Figure 4.7). Those were potential points of interest, which needed further clarification in order to obtain a better introspective insight into the nature of the CALL qualities under the investigation. The interview questions that originated from the observation notes on Student 28 in the example above are provided in Figure 4.8. At one instance during the observation of this student, having heard in the think-aloud that the student noticed a mismatch between the color-coded feedback and the move he intended to construct, the researcher made an entry in her observation notes along with a question mark, based on which during the interview she asked the student about his thoughts at that time.

Researcher: What did you think about when you noticed that the move you had in mind was displayed in a different color by the program?
…
Researcher: At that point, after having thought a little bit about that sentence, you said that the program was right. What made you say that?
…
Researcher: What influenced your decision to revise different pieces of your draft?
…
Researcher: It seemed to me that you took time to think quite frequently. Am I right? What were you mostly thinking about? (Student 28, interview)

Figure 4.8. Examples of interview questions

Audacity, a free open source digital audio editor and recording application was used to record the interviews audio data, which was saved as .wav files suitable for importing audio data into Transana. Although Camtasia could have been used to record the interview data as well, Audacity was preferred because the size of the files it generates is much smaller compared to those of Camtasia.

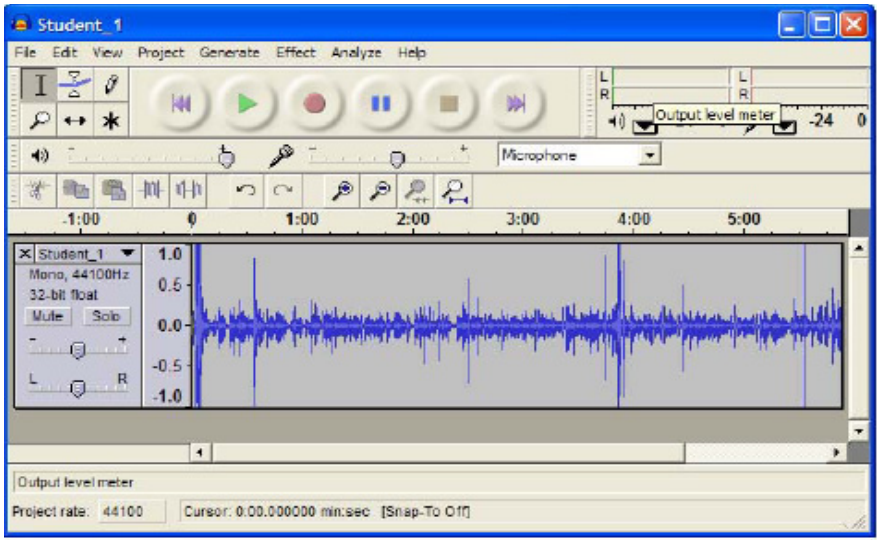


Figure 4.9. Screenshot of Audacity

4.3.8. Transana

Transana, an open source qualitative data analysis software developed at the Wisconsin Center for Education Research at the University of Wisconsin-Madison, was used for transcription and coding. The data transcribed and coded in Transana were the interviews, Camtasia screen recordings, and think-aloud protocols. The observations transcripts were also imported into Transana. These data were introduced in Transana in so-called ‘series’, which, for each student, contained an interview audio and a transcript, an imported transcript of the observation, and a video episode of Camtasia screen capturing with the accompanying think-aloud (Figure 4.10). The last two data sources were combined in one transcript, covering the student’s utterances as well as the accompanying actions as captured on the screen. Camtasia also made it possible to include in the transcript the specific modifications made by the students during revision with IADE. Such an example is given in Figure 4.11, where, in addition to the student’s words, there are researcher’s comments in square brackets and text changes in quotation marks.

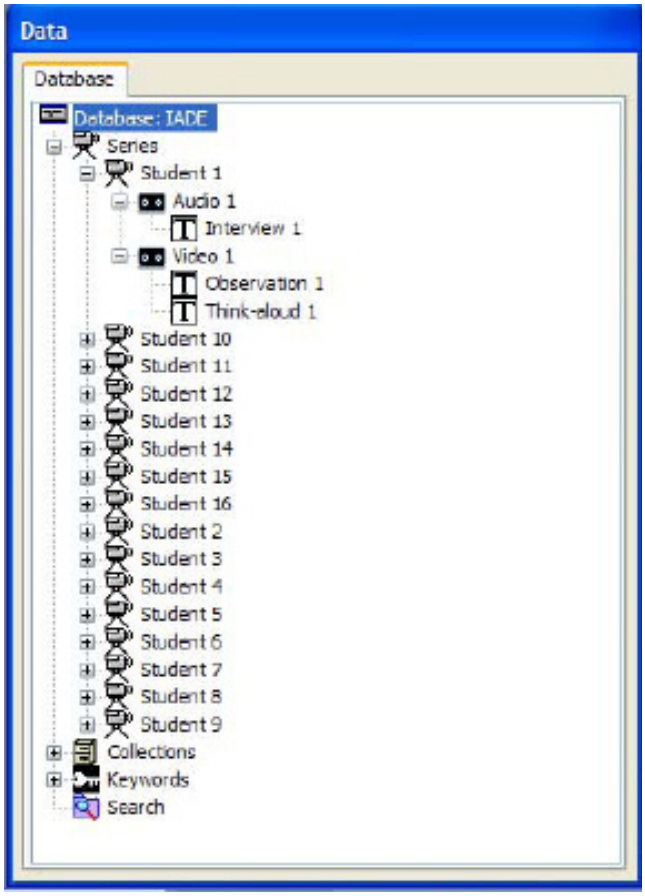


Figure 4.10. Data organization in Transana

¤<1659145> [reads and thinks] I'm still looking at the last sentence ... and to see if I can make some changes and the program can respond to these changes. Maybe I can try to be specific and say that "the results of this paper". Yeah, let's try to do that. [makes the change: "The results show that..." → "The results of this study show that..." and submits]

Figure 4.11. Excerpt from Student 27 transcript in Transana

4.3.9. Coding taxonomy

The analysis of qualitative data was done according to a coding taxonomy developed based on the results of the pilot study conducted in July 2008. The taxonomy consisted of coding categories for the four CALL qualities investigated: Language Learning Potential, Meaning Focus, Learner Fit, and Impact. Each of these four categories included a number of

sub-categories shown in Figures 4.12a and further exemplified in Section 4.6. These sub-categories emerged from the 2008 pilot study data such as think-aloud protocols, interviews, and observations. A second coder was not involved since that would require extensive training as well as deep theoretical knowledge. To ensure the reliability of coding, the researcher coded four think-aloud protocols, four interviews, and four observations from the pilot study data twice with an interval of eight months, first in August 2008 and second in April 2009. The comparison of coded data helped confirm and refine the initial coding categories, some of which were combined as shown in Table 4.2. The revised categories we used in coding the data for this study.

Table 4.2. Initial and revised coding categories

Initial coding categories	*Revised coding categories*
Language Learning Potential – focus on discourse form – focus on feedback – focus on length of discourse components – noticing of negative evidence – output modification – understanding of discourse moves and their functions	Language Learning Potential – focus on discourse form – noticing of negative evidence – output modification – enhanced understanding
Meaning Focus – focus on discourse meaning – identifying a connection bw meaning and lexical choice – noticing of mismatch bw intended meaning and feedback – reflecting on meaning	Meaning Focus – noticing a mismatch – reflection on functional meaning – connection bw functional meaning and lexical choice – construction of new functional meaning
Learner Fit – appraisal of evaluation method – feedback appropriate for learner characteristics – help options – helpful feedback – level of difficulty – learner control – practice opportunity	Learner Fit – appropriateness of evaluation method – appropriateness of feedback for task completion – facilitating problem-solving – fit for individual learner characteristics – learner control

Table 4.2. Initial and revised coding categories (continued)

Initial coding categories	*Revised coding categories*
Impact – cognitive involvement – emotional involvement – influence on the revision process – motivation – negative learning experience – positive learning experience – strategy development – desire to use the program	Impact – cognitive involvement – emotional involvement – influence on the revision process – motivation – negative learning experience – positive learning experience – strategy development

4.4. Procedure

IADE was implemented as a part of classroom instruction during the Fall semester of 2008 and the Spring semester of 2009. Three groups of students participated in this study in each of these two semesters, the numbers being almost the same – 53 students in Fall 2008 and 52 students in Spring 2009. According to the course syllabus, the students were taught the writing conventions of research article introductions in class based on Swales' (1981, 1990, 2004) move schema. One class period was devoted to studying each move; for that purpose, the students worked with their corpora as described in Section 3.1.3. Then, the students were given the pre-test on the introduction moves and steps. The pre-test was delivered at this time as opposed to prior to instruction because the purpose was to measure not the learning gains after instruction, but rather after revising with IADE. After the pre-test, the students were required to write a draft of their own introduction sections for their research articles as homework.

The next class period, the instructor introduced IADE, explained the way it functioned, and demonstrated how to work with it. It was also explained to the students that their task was to revise and make improvements on their draft. The interaction consisted of

submitting the draft for automated evaluation, receiving immediate individualized feedback, making revisions, and resubmitting the new draft to the system. This was an iterative process, which began in class and ended outside of class, and that allowed the students to spend as much time on practice and revision as they needed. During the following class session, when the final introduction draft was due for submission, the post-test was given. After the post-test, the survey questions were delivered. They were completed and returned by the participants to the researcher in electronic form at the time of their best convenience within a week.

The value of these data for the purpose of this research was explained in participants' consent forms, which were handed out prior to data collection. The ISU Institutional Review Board approved two consent forms at the same time – one for all the participants in the study, and the second for students who would volunteer to participate in additional data collection through computer screen recording, think-aloud protocols, observations, and interviews. Volunteers were recruited only in the Fall 2008 semester. Sixteen participants signed the second consent form and therefore completed the revision task with IADE in an experimental computer lab setting under observable conditions and in the presence of the researcher. Three of these students interacted with IADE after they had begun revising with the program in class, and the other 13 students worked with the program before it was introduced in class, which is why they were first provided with the same kind of tutorial information about the goals and the functionality of IADE as all the other participants. The other 37 Fall semester participants, as well as the 52 Spring semester participants, signed the first consent form, in which they agreed to revise their introduction drafts with IADE, to complete pre- and post-

tests, and to answer survey and scale questions. The different procedures for the two semesters are summarized in Figures 4.12a and 4.12b.

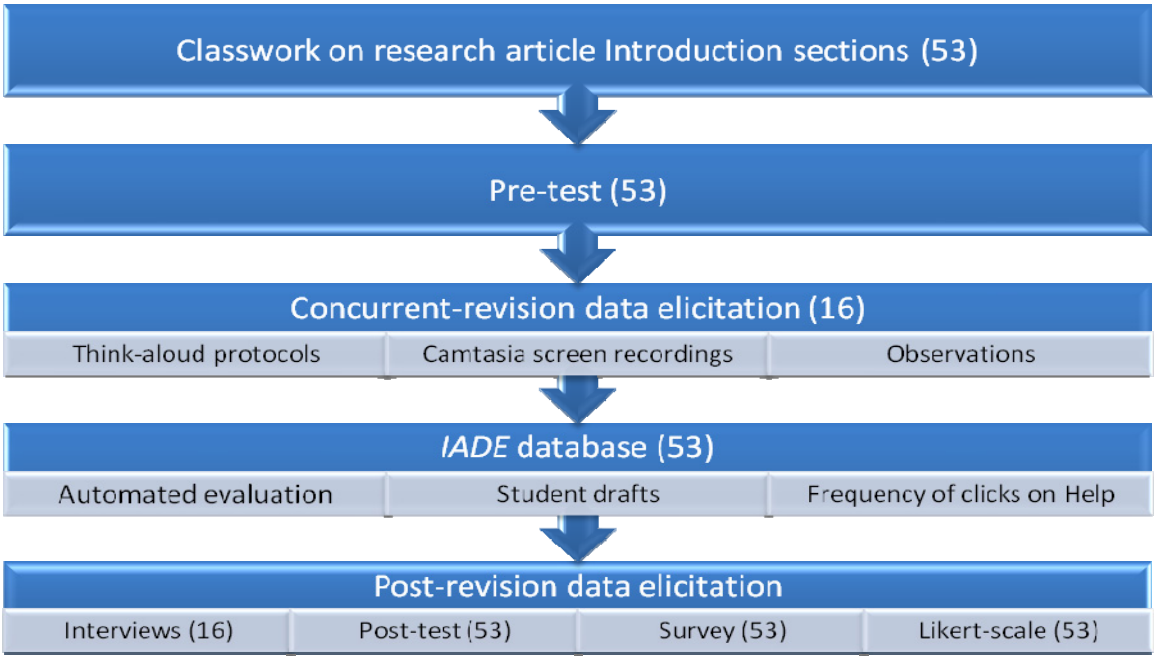


Figure 4.12a. Procedure for the Fall 2008 semester

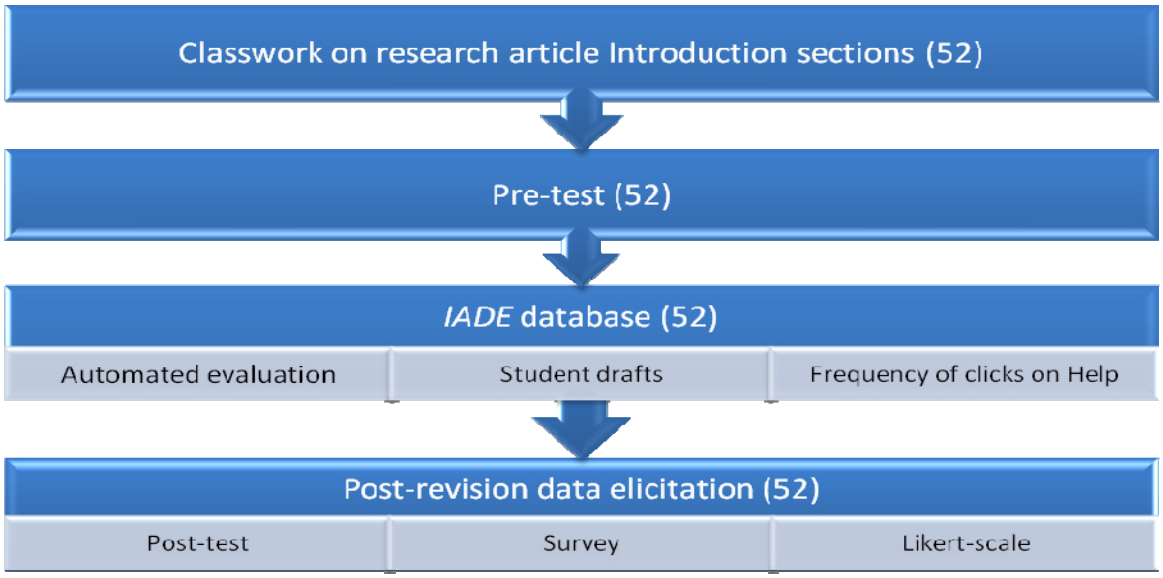


Figure 4.12b. Procedure for the Spring 2009 semester

4.5. Data

The types and the amount of collected data are shown in Table 4.3. It should be clarified here that the 16 first and 16 last drafts analyzed manually were written by the 16 participants who volunteered to use IADE in an experimental setting. The writing of these particular participants was chosen so that the results of the analysis could be triangulated with their think-aloud, observation, and interview data in order to provide a better understanding of uncovered phenomena. Also, the same drafts were scored by human raters, who had no knowledge of which drafts were first and last since the documents contained only the text and were named in a way that only the researcher could decode.

Table 4.3. Summary of data

Data	*N. participants*	*Extent*
1. Pre-test	104	2 tasks (17 sentences each)
2. Post-test	104	2 tasks (17 sentences each)
3. Camtasia recordings 4. Think-aloud protocols	16 16	Student 1 (ID#27) – 29 minutes Student 2 (ID#29) – 36 minutes Student 3 (ID#28) – 37 minutes Student 4 (ID#30) – 29 minutes Student 5 (ID#32) – 28 minutes Student 6 (ID#43) – 31 minutes Student 7 (ID#44) – 26 minutes Student 8 (ID#53) – 38 minutes Student 9 (ID#40) – 30 minutes Student 10 (ID#54) – 27 minutes Student 11 (ID#58) – 30 minutes Student 12 (ID#61) – 32 minutes Student 13 (ID#62) – 25 minutes Student 14 (ID#63) – 26 minutes Student 15 (ID#64) – 32 minutes Student 16 (ID#65) – 30 minutes
5. Observations	16	1-2 pages (12, Times New Roman, single)
6. Interviews	16	10 – 15 minutes
7. Yes/No and open-ended survey questions	83	25 questions
8. Likert-scale survey questions	88	19 questions

Table 4.3. Summary of data (continued)

Data	*N. participants*	*Extent*
9. IADE's evaluation of first and last drafts	105	% and comment comparison to discipline for 210 drafts
10. Nr of clicks on Help Options	105	total of 1312
11. Students' first drafts	105	16 drafts for manual analysis
12. Students' last drafts	105	16 drafts for manual analysis

The general approach of the mixed methods concurrent transformative strategy presented in Figure 4.1 is diagrammed in terms of the specifics of this study in Figure 4.13. Under the umbrella of the CALL evaluation conceptual framework by Chapelle (2001), the quantitative and qualitative data, enumerated in the first two oval shapes, were collected concurrently and integrated during the analysis and interpretation stage, shown in the circle on the right. At this stage, the focus was on quantitative evidence, which was collected from 105 participants. Qualitative data, consisting of 83 participants' open-ended survey responses as well as audio and screen recordings of 16 participants, were examined to obtain a deeper understanding of the investigated phenomena and to support the generalizing claims to be made based on numerical results.

Both quantitative and qualitative data were analyzed in order to answer each research question for the four CALL qualities outlined in the guiding framework adopted in this study. Table 4.4 summarizes the types of data examined in seeking evidence for each quality. Further, the contents of this table are explained in terms of conducted analyses.

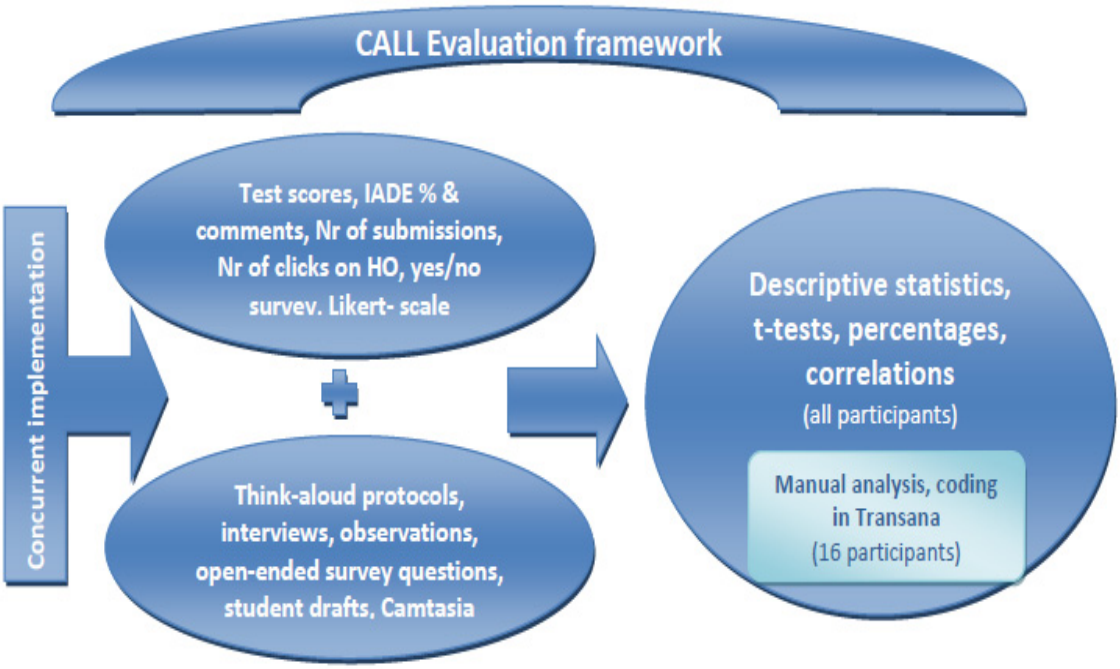


Figure 4.13. The concurrent transformative strategy applied in this study

Table 4.4. Data sources for CALL qualities

CALL quality	*Data source*
Language learning potential	*Quantitative:* yes/no responses, Likert-scale responses, pre/post-test scores, IADE's automated evaluation, number of submissions, number of clicks on Help Options *Qualitative:* open-ended survey responses, think-aloud protocols, Camtasia recordings, observations, semi-structured interviews, text analysis of first and last drafts, human rating of first and last drafts
Meaning focus	*Quantitative:* yes/no responses, Likert-scale responses *Qualitative:* open-ended survey responses, think-aloud protocols, Camtasia recordings, observations, semi-structured interviews
Learner fit	*Quantitative:* yes/no responses, Likert-scale responses *Qualitative:* open-ended survey responses, think-aloud protocols, Camtasia recordings, observations, semi-structured interviews
Impact	*Quantitative:* yes/no responses, Likert-scale responses *Qualitative:* open-ended survey responses, think-aloud protocols, Camtasia recordings, observations, semi-structured interviews

4.6. Data Analysis

4.6.1. Language Learning Potential

The investigation of Language Learning Potential relied on analyses of both quantitative and qualitative data. The research questions posed inquired whether learners' interaction with IADE would result in students' noticing of and focus on discourse form and in remediation indicative of learning of the discourse forms. In addition to the survey data, the latter was examined from the perspectives of rhetorical improvement in students' writing as reflected in their first and last drafts as well as their learning gains as reflected in their pre/post-test scores. Table 4.5 summarizes the data sources that yielded evidence for the aspects of Language Learning Potential.

Table 4.5. Aspects, data, and analyses of Language Learning Potential

Aspects of LLP	*Data and analysis*	
Noticing and focus on discourse form	*Quantitative*: yes/no responses; Likert-scale responses	comparison of response percentages
	Qualitative: open-ended survey responses; think-aloud protocols; observations; semi-structured interviews	manual analysis of responses; coding
Improvement in the rhetorical quality of writing	*Quantitative*: yes/no responses; Likert-scale responses; automated evaluation of first and last drafts converted to scores	comparison of response percentages; descriptive statistics of first and last draft scores for moves and length; t-tests for first and last draft scores
	Qualitative: open-ended survey responses; think-aloud/Camtasia transcripts; student first and last drafts	manual analysis of responses; coding; manual analysis of output modifications in 16 first and 16 last drafts
Learning gains	*Quantitative*: yes/no responses; Likert-scale responses; pre- and post-test scores	comparison of response percentages; descriptive statistics and t-tests for pre- and post-test scores
	Qualitative: open-ended survey responses	manual analysis of responses

Table 4.5. Aspects, data, and analyses of Language Learning Potential (continued)

Aspects of LLP	*Data and analysis*	
Practice	*Quantitative*: yes/no responses; Likert-scale question; number of submissions; automated evaluation of first and last drafts converted to scores	comparison of response percentages; Pearson product-moment correlation; One-Way ANOVA
	Qualitative: open-ended survey responses	manual analysis of responses
Modified interaction	*Quantitative*: yes/no responses; Likert-scale responses; number of clicks on Help Options; automated evaluation of first and last drafts converted to scores	comparison of percentages for frequencies of access to Help Options; Pearson product-moment correlation; One-Way ANOVA
	Qualitative: open-ended survey responses; think-aloud/Camtasia transcripts	manual analysis of responses; coding

4.6.1.1. Noticing of and focus on discourse form

Both quantitative and qualitative data were analyzed and integrated to obtain evidence of students' focus on discourse form. Percentages for yes/no and Likert-scale responses covering this point of interest were calculated and compared. Deeper introspective insights were then gained from the open-ended survey responses, think-aloud protocols, semi-structured interviews, and observations. The latter three data sources were analyzed in Transana according to the revised coding categories introduced in Section 4.3.9. Following are examples for each of the Language Learning Potential coding categories:

- Focus on discourse form

  I find many colors mixing around; it's not very clear structure. (Student 54, think-aloud protocol transcript)

  The feedback helps me organize the distribution of the moves in introductions. And it also helps me pay more attention on the moves which need to be focused on. (Student 27, semi-structured interview transcript)

Colors appear to really help him focus on the moves. (Student 53, observation transcript)

- Noticing of negative evidence

Actually… I want to start my draft with move 1, but the software tells me that I started with move 2. (Student 27, think-aloud protocol transcript)

So, I paid attention to that only because the program showed me that move 2 is missing. (Student 28, semi-structured interview transcript)

He thinks back to the way he thought he organized the structure of his introduction and, looking at the color-coded feedback, notices that the cyclical approach to developing the introduction that he wanted to take was not really displayed in cycles by the program. (Student 32, observation transcript)

- Output modification

Let me try this sentence. [goes to the revision box. reads a sentence from there] maybe I can say like this - "we chose the wavelength from 532 nanometers" [makes this change] ok. [submits] (Student 40, think-aloud protocol transcript)

Wants to express m2 Hypothesis and changes "can" to "could". (Student 44, observation transcript)

I changed some expressions, especially when I see that is not the right color. (Student 64, semi-structured interview transcript)

- Enhanced understanding

Oh, I made a statement which justifies the fact, but I don't know why it didn't show up as move 2. [goes back to class materials and reads about justification] Oh, I see. … So I guess this is a sort of justifying the bigger statement, like why this is so. So that's a different kind of justification. It's not what I thought. (Student 54, think-aloud protocol transcript)

I think the program makes you understand theory better. For example, for me wasn't enough to just read the explanation of theory. […] But when you actually write something and get response on how you did, it helps better understand the moves and steps, and to see that I may understand something wrong. (Student 30, semi-structured interview transcript)

At this point he realizes that maybe he doesn't understand the 'justification' properly. Therefore he checks the class materials again trying to understand the 'justification' from the examples provided there. He seems to understand now the

real meaning of the 'justification' step and decides that he can still justify why it is worth addressing this particular topic." (Student 61, observation transcript)

4.6.1.2. Rhetorical improvement

In the context of this study, the rhetorical competence to be developed covered the ability to effectively organize a research article introduction, operating with all the three moves described by Swales (1981, 1990, 2004), and to render their functional meaning through the use of appropriate steps. The specific communicative goals the students had to achieve in their writing were:

- To demonstrate the centrality of their topic in the field
- To demonstrate knowledge on the selected research topic
- To argue for the need to conduct more research
- To demonstrate how the reported study fulfills that need
- If practiced by the discipline: (a) to demonstrate the value of the reported study, (b) to preview the content of the article

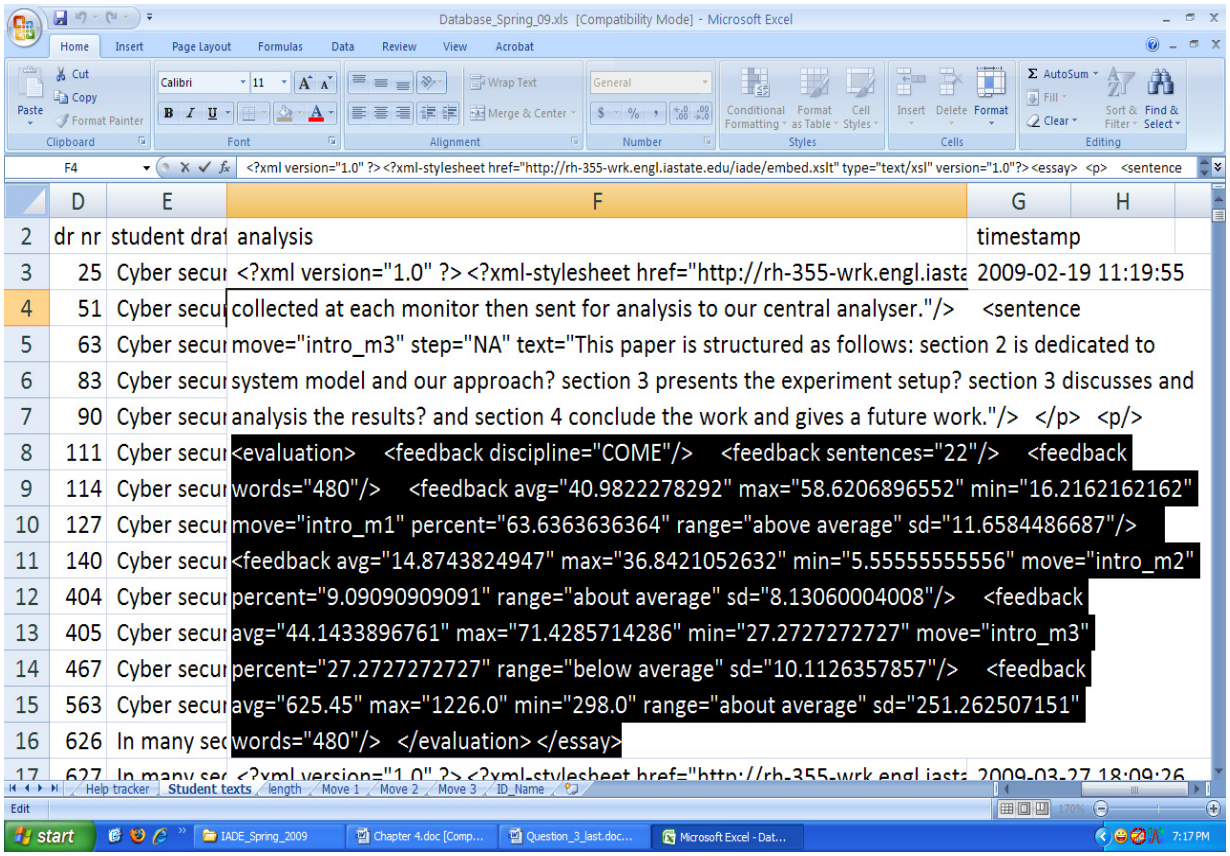
In order to find whether the students' writing, and more specifically, the rhetorical effect of their introductions, indeed improved after their revision using IADE, the study employed a protocol that included analyses of automatically generated scores as well as of holistic and analytic human ratings for first and last drafts (see Table 4.6).

Table 4.6. Data analysis for improvement in rhetorical competence

Evidence sought	*Data source*	*Analysis*
Improvement in discourse form (i.e., distribution of moves and changes in length)	IADE's automated evaluation of 105 first and 105 last drafts	descriptive statistics, t-tests for first and last draft moves and length
Improvement in discourse quality (i.e., presence and effectiveness of moves)	Human rating of 16 first and 16 last drafts	descriptive statistics, t-tests for first and last draft moves and length; manual analysis of output modifications

IADE evaluated student drafts by returning percentages along with evaluative comments for each move, which represented the distribution of the moves in the draft compared to that in the students' discipline. It also returned the number of words along with comments similar to the ones provided for each move. This kind of automated evaluation was stored and accessed in the database in the form displayed in Figure 4.14, where the automated evaluation of the draft is highlighted.

As it can be seen in Figure 4.14, the evaluation part begins with the notation <evaluation>, which is followed by the abbreviation of the name of the discipline to which the student's draft was compared and based on which the feedback was provided (which is why the word 'feedback' appears there as well). In this example, COME stands for Computer Engineering. Following is a count of the sentences and of the words in the draft, after which the average, maximum, and minimun of words in this discipline are given (i.e., <feedback avg="40.9822278292" max="58.6206896552" min="16, 2162162162" respectively). Next, is the evaluation of Move 1 (move="intro_m1" percent=63. 6363636364" range="above average"). The 'above average' comment means that the student's draft has more Move 1 compared to the average in the Computer Engineering discipline. The average, maximum, and minimum percentages for this move in this particular discipline are also included (<feedback avg="14.8743824947" max="36.8421052632" min="5.55555555556") so that the feedback helps the student realize how far or how close s/he is from the norm. The evaluation of Move 2 and Move 3 is the same as that of Move 1. It should be noted here that the average, maximum, and minimum figures for the introductions of a certain discipline result from a preliminary analysis of the corpus in that discipline (here COME), which was used in the training of IADE's analysis module.



	D	E	F	G	H
2	dr nr	student drat	analysis	timestamp	
3	25	Cyber secur	<?xml version="1.0" ?> <?xml-stylesheet href="http://rh-355-wrk.engl.iasta	2009-02-19 11:19:55	
4	51	Cyber secur	collected at each monitor then sent for analysis to our central analyser."/> <sentence		
5	63	Cyber secur	move="intro_m3" step="NA" text="This paper is structured as follows: section 2 is dedicated to		
6	83	Cyber secur	system model and our approach? section 3 presents the experiment setup? section 3 discusses and		
7	90	Cyber secur	analysis the results? and section 4 conclude the work and gives a future work."/>		
8	111	Cyber secur	<evaluation> <feedback discipline="COME"/> <feedback sentences="22"/> <feedback		
9	114	Cyber secur	words="480"/> <feedback avg="40.9822278292" max="58.6206896552" min="16.2162162162"		
10	127	Cyber secur	move="intro_m1" percent="63.6363636364" range="above average" sd="11.6584486687"/>		
11	140	Cyber secur	<feedback avg="14.8743824947" max="36.8421052632" min="5.55555555556" move="intro_m2"		
12	404	Cyber secur	percent="9.09090909091" range="about average" sd="8.13060004008"/> <feedback		
13	405	Cyber secur	avg="44.1433896761" max="71.4285714286" min="27.2727272727" move="intro_m3"		
14	467	Cyber secur	percent="27.2727272727" range="below average" sd="10.1126357857"/> <feedback		
15	563	Cyber secur	avg="625.45" max="1226.0" min="298.0" range="about average" sd="251.262507151"		
16	626	In many sec	words="480"/> </evaluation> </essay>		
17	627	In many sec	<?xml version="1.0" ?> <?xml-stylesheet href="http://rh-355-wrk.engl.iasta	2009-03-27 18:09:26	

Figure 4.14. Example of IADE's automated evaluation stored in the database

From all this information, only the evaluative comments were used in the analysis. Because the program did not give scores, the comments in the feedback were used to assign the following scores, which helped determine improvement from first to last draft:

IADE comment: 'about average' → score: 3

IADE comment: 'below average' or 'above average' → score: 2

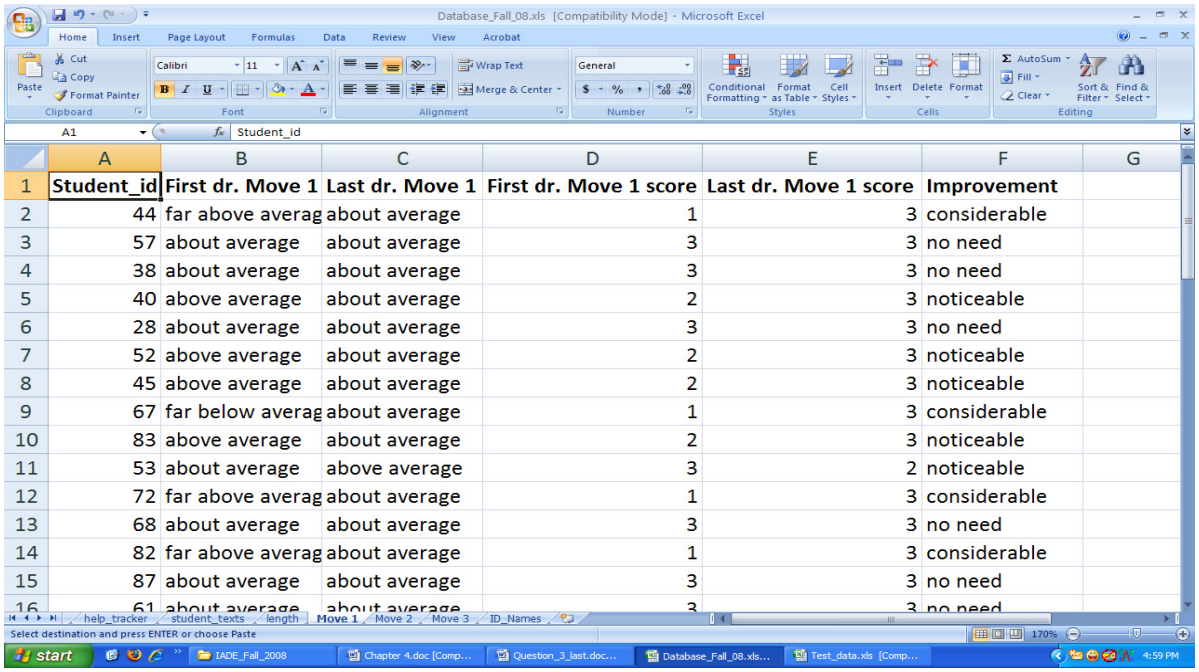
IADE comment: 'way below average' or 'way above average' → score: 1

Then, descriptive statistics were calculated and the means for Move 1, Move 2, Move 3, and drafts' length were compared through t-tests to find if the improvement in all these aspects was statistically significant.

To obtain a better understanding of the degree of rhetorical improvement, improvement was classified into four categories based on the assigned scores (see Table 4.7 and Figure 4.15). It should be noted that in the few cases when the first draft received a comment in the "average" range, qualifying for a score of 3, and the last draft received a comment in the "above" range, qualifying for a score of 2, the improvement was also considered noticeable. When the first draft received a comment in the "below" range and the last draft received a comment in the "above" range, the improvement was regarded as considerable.

Table 4.7. Categories of improvement in rhetorical moves and draft length

First draft scores	*Last draft scores*		*Improvement category*
1	3	→	Considerable improvement
1, 2	2, 3 (respectively)	→	Noticeable improvement
1, 2 or 3, 2	1, 2 or 2,1 (respectively)	→	No improvement
3	3	→	No need for improvement



Student_id	First dr. Move 1	Last dr. Move 1	First dr. Move 1 score	Last dr. Move 1 score	Improvement
44	far above averag	about average	1	3	considerable
57	about average	about average	3	3	no need
38	about average	about average	3	3	no need
40	above average	about average	2	3	noticeable
28	about average	about average	3	3	no need
52	above average	about average	2	3	noticeable
45	above average	about average	2	3	noticeable
67	far below averag	about average	1	3	considerable
83	above average	about average	2	3	noticeable
53	about average	above average	3	2	noticeable
72	far above averag	about average	1	3	considerable
68	about average	about average	3	3	no need
82	far above averag	about average	1	3	considerable
87	about average	about average	3	3	no need
61	about average	about average	3	3	no need

Figure 4.15. Excerpt from data on improvement in Move 1

The analysis was then carried out as indicated in Table 4.8 where percentages are given along with the number of students who improved at each of the four levels.

Table 4.8. Degrees of improvement in rhetorical moves and text length

Improvement category	*Move 1*	*Move 2*	*Move 3*	*Length*	*Average improvement*
Considerable improvement	% (n)	% (n)	% (n)	% (n)	% (n)
Noticeable improvement	% (n)	% (n)	% (n)	% (n)	% (n)
No improvement	% (n)	% (n)	% (n)	% (n)	% (n)
No need for improvement	% (n)	% (n)	% (n)	% (n)	% (n)

To find whether improvement also occurred in the rhetorical quality of the submitted introductions, the first and the last drafts of the 16 volunteering students were manually analyzed and rated. Scores of 1, 2, and 3 were assigned to maintain similarity with the scores based on the program's automated evaluation. (The rating procedure as well as the reliability among raters and human-IADE reliability are presented further in section 4.6.5.)

To be consistent with the improvement analysis based on IADE's automated evaluation, descriptive statistics and t-tests with the scores assigned by human raters to the 16 students' moves in first and last draft were also calculated. Then, the scores were classified into the same four improvement categories: considerable, noticeable, no, and no need for improvement.

In addition, the output modifications made by the same 16 participants from first to last draft were manually analyzed. They were identified in Camtasia screen recordings as well as in a direct comparison of first and last drafts. The modifications were categorized into six levels: content, lexical, grammatical, structural, and mechanics. Each type of modification was then quantified in percentages.

4.6.1.3. Effect of practice on improvement

One of the purposes of IADE's implementation was to provide students with sufficient writing practice through multiple draft revisions and iterative resubmissions. Consequently, determining whether such practice had an effect on rhetorical improvement was expected to yield additional evidence of learning potential. Submission frequency data was extracted from IADE's database, which stored the number of drafts submitted by every student and the time of submission, and was used as an indicator of practice. Improvement was reflected in the score difference between first and last drafts.

The effect of practice on the improvement of Move 1, Move 2, Move 3, and length was explored through One-Way Analysis of Variance (ANOVA). The number of draft submissions was considered as the independent variable. The score differences for each move and length between first and last drafts were considered to be the dependent variables. Considering that the assigned scores were 1, 2, and 3, the score difference fell into one of the four groups shown in Table 4.9 below. The last group (improvement = -1) had only one observation for Move 1 and was, thus, excluded from the analysis. It must be noted here that a lower score for that move in the last draft does not necessarily mean that there was a regress in its quality since it was due to an increase in the length of that move; IADE's comment on it was "above average" compared to the norms of the discipline. ANOVA was used to compare the means of the remaining three groups in order to see if the number of submissions was different for improvement at different levels, for a module at a time (i.e., for Move 1, Move 2, Move 3, and length).

Table 4.9. Score difference as the improvement variable

First draft scores	*Last draft scores*	*Score difference*	*Improvement*
1	3	2	highest
1, 2	2, 3 (respectively)	1	moderate
1, 2, 3	1, 2, 3 (respectively)	0	no improvement
3, 2	2, 1 (respectively)	-1	"regress"

The effect of practice on the overall draft improvement was examined with the help of the Pearson product-moment correlation. The same number of draft submissions used in the ANOVA analyses was correlated with the sum of score differences for each move and length, which was used as an indicator of overall draft improvement. For example, if the score difference for a student's Move 1 was 2, for Move 2 was 1, for Move 3 was 2, and for length was 0, then the overall draft improvement was the sum of these figures, amounting to 5. This number was then correlated with the number of submissions this particular student made.

4.6.1.4. Learning gains

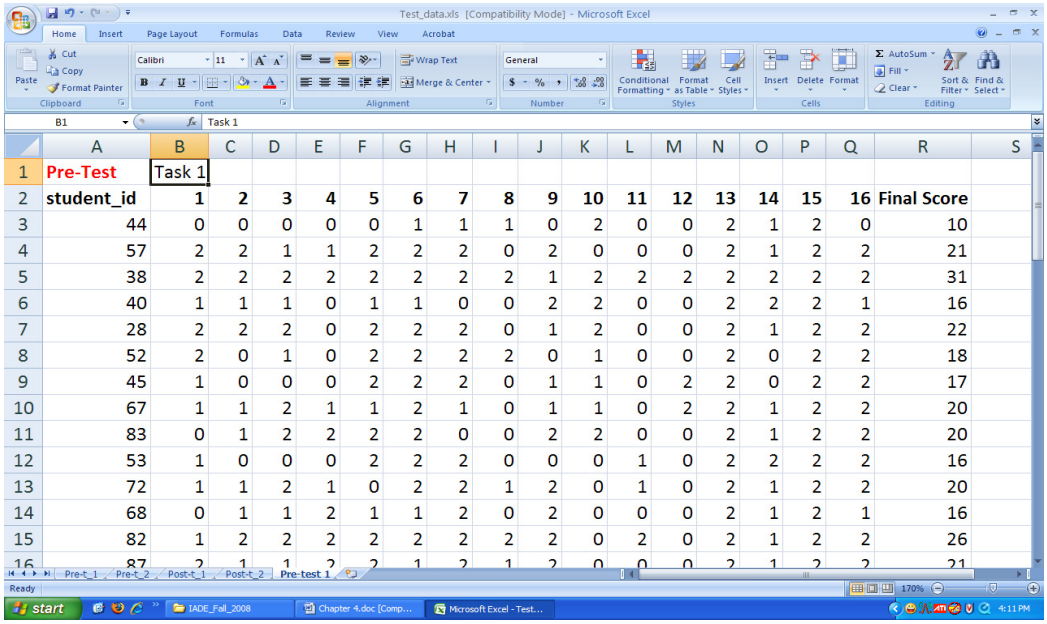
The pre- and post-tests were scored for every sentence in each of the two tasks since the tasks required the students to identify the moves and the steps represented by the given sentences. The individual sentences in task one had unambiguous and clearly expressed functions realized through vocabulary that signaled a certain move and step. Similarly, the texts chosen for the second task had a clear rhetorical development signposted by functional lexical items. Only one correct answer was possible for each sentence in both test tasks. A score of 2 was assigned for a correct move and a correct step; 1 – for an incorrect move, but a correct step, or for a correct move, but an incorrect step; and 0 – for an incorrect move and an incorrect step (see Table 4.10).

Table 4.10. Scoring of pre- and post-tests

Student answer	*Score*
correct move/correct step	2
incorrect move/correct step	1
correct move/incorrect step	1
incorrect move/incorrect step	0

Descriptive statistics and t-tests were calculated for each task. Then, the overall individual scores obtained by summing up the scores for each question in the test (Figure 4.16) were used to calculate descriptive statistics and to run another t-test. This allowed for the comparison of means representing students' knowledge of moves before and after the interaction with IADE.

The ways in which the students perceived their learning of moves and writing improvement were elucidated additionally with the help of their Likert-scale and yes/no answers, for which percentages were calculated, and with the help of open-ended responses, which were examined for the presence of emerging themes.

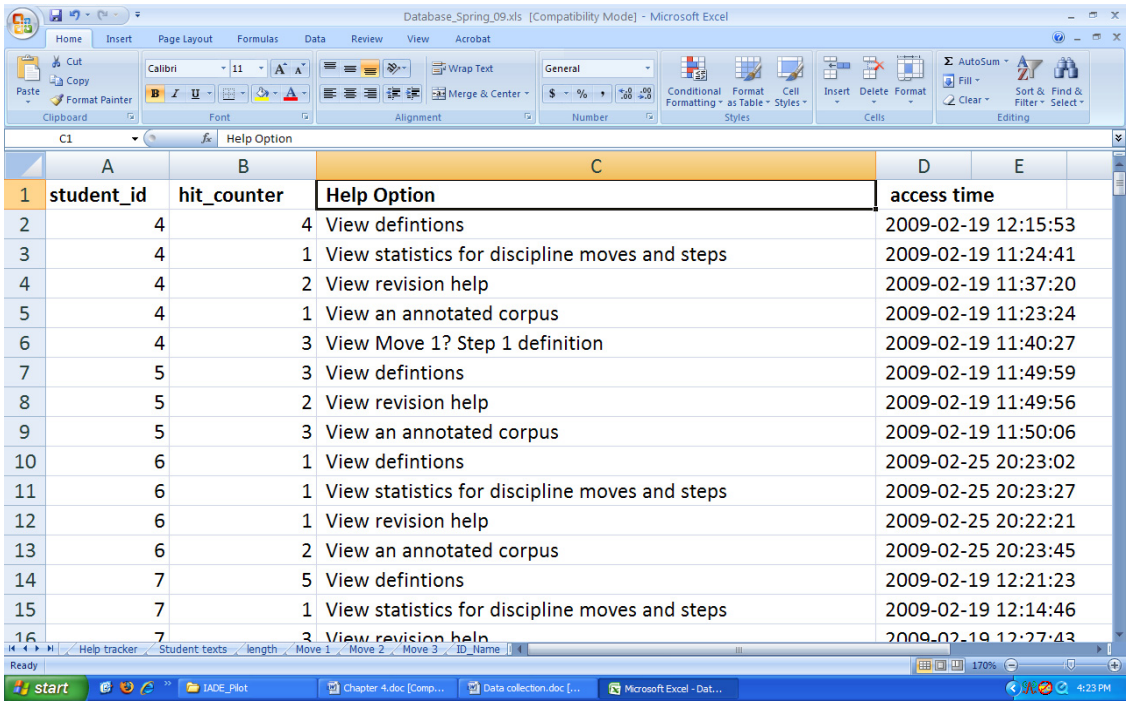


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Pre-Test	Task 1																
2	student_id	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Final Score
3	44	0	0	0	0	0	1	1	1	0	2	0	0	2	1	2	0	10
4	57	2	2	1	1	2	2	2	0	2	0	0	0	2	1	2	2	21
5	38	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	31
6	40	1	1	1	0	1	1	0	0	2	2	0	0	2	2	2	1	16
7	28	2	2	2	0	2	2	2	0	1	2	0	0	2	1	2	2	22
8	52	2	0	1	0	2	2	2	2	0	1	0	0	2	0	2	2	18
9	45	1	0	0	0	2	2	2	0	1	1	0	2	2	0	2	2	17
10	67	1	1	2	1	1	2	1	0	1	1	0	2	2	1	2	2	20
11	83	0	1	2	2	2	2	0	0	2	2	0	0	2	1	2	2	20
12	53	1	0	0	0	2	2	2	0	0	0	1	0	2	2	2	2	16
13	72	1	1	2	1	0	2	2	1	2	0	1	0	2	1	2	2	20
14	68	0	1	1	2	1	1	2	0	2	0	0	0	2	1	2	1	16
15	82	1	2	2	2	2	2	2	2	2	0	2	0	2	1	2	2	26
16	87	2	1	1	2	2	1	2	1	2	0	0	0	2	1	2	2	21

Figure 4.16. Excerpt from test scores data

4.6.1.5. Effect of modified interaction on improvement

Another aspect of IADE's learning potential, modified interaction facilitated by its Help Options, was analyzed with the help of automatically tracked students' clicks on the four available options: Definitions of moves and steps, Step Statistics, Annotated Corpus, and Revision Tips (see Figures 4.4e and 4.17 for excerpts from the data).



student_id	hit_counter	Help Option	access time
4	4	View defintions	2009-02-19 12:15:53
4	1	View statistics for discipline moves and steps	2009-02-19 11:24:41
4	2	View revision help	2009-02-19 11:37:20
4	1	View an annotated corpus	2009-02-19 11:23:24
4	3	View Move 1? Step 1 definition	2009-02-19 11:40:27
5	3	View defintions	2009-02-19 11:49:59
5	2	View revision help	2009-02-19 11:49:56
5	3	View an annotated corpus	2009-02-19 11:50:06
6	1	View defintions	2009-02-25 20:23:02
6	1	View statistics for discipline moves and steps	2009-02-25 20:23:27
6	1	View revision help	2009-02-25 20:22:21
6	2	View an annotated corpus	2009-02-25 20:23:45
7	5	View defintions	2009-02-19 12:21:23
7	1	View statistics for discipline moves and steps	2009-02-19 12:14:46
7	3	View revision help	2009-02-19 12:27:43

Figure 4.17. Excerpt from Help Options access data (this is raw data that was then transferred to a separate worksheet shown in Figure 4.4e)

To establish whether modified interaction had an effect on writing improvement, the same kinds of statistical analyses as for the effect of practice on improvement were done (see Section 4.6.1.3 above) – One-Way ANOVA for improvement on individual aspects of the drafts and Pearson for overall draft improvement. The variables used in this analysis were the frequency of students' access to each of the four available Help Options and the difference between the first and last draft scores for the three moves.

4.6.2. Meaning Focus

To evaluate IADE's Meaning Focus quality, two aspects were examined: focus on and construction of functional meaning. The quantitative and qualitative data enumerated in Table 4.11 were analyzed to find whether feedback could enhance focus on discourse meaning and whether this could lead to successful construction of such meaning.

Table 4.11. Aspects, data, and analyses of Meaning Focus

Aspect of MF	*Data and analysis*	
Focus on functional meaning of discourse	*Quantitative*: yes/no responses; Likert-scale responses	comparison of response percentages
	Qualitative: open-ended survey responses; think-aloud protocols; observations; semi-structured interviews	manual analysis of responses; coding
Construction of functional meaning of discourse	*Qualitative*: think-aloud protocols; Camtasia screen recordings; observations; semi-structured interviews; open-ended survey responses	manual analysis of responses; coding

Student responses to 3 survey and 3 Likert-scale questions were examined and percentages for those were calculated. Further, whether students focused on meaning upon the reception of feedback was examined with the help of qualitative data from think-aloud protocols, observations, semi-structured interviews, and open-ended survey responses. With the exception of the latter, where emerging themes could not be predicted, all the recorded and transcribed data were coded in Transana in terms of the revised taxonomy categories (see Section 4.3.9).

- Reflection on functional meaning

  [reads from draft] …"becomes a key question" OK, good, this should be move 2 because I want to raise a question, it is what I mean." (Student 58, think-aloud protocol transcript)

Seems to think that what he really wants to express is the advantage of his method. (Student 43, observation transcript)

"I think in this case I started to think more about the meaning of my sentence…" (Student 40, interview)

- Noticing of mismatch between feedback and intended meaning

The second sentence, I think it should be blue. That means it should be move 1, but it turned out to be move 2. Maybe the sentence is not clearly or correctly written. (Student 43, 1think-aloud protocol transcript)

He still has a problem. Part of m3 appears as m1. (Student 43, observation transcript)

The program makes me notice that what I mean is not what I say. (Student 29, interview transcript)

- Identifying a connection between meaning and lexical choice

I think I want to express the advantage of our method; it should be m3. And, uh… [reads] "There's no need" Maybe this "no" is misleading." (Student 58, think-aloud protocol transcript)

He spots the word "past", thinking that it is the reason for why his intended m3 appears as m2. (Student 65, observation transcript)

I tried to see some examples and expressions, and sometimes I could find some word that fit very well with what I want to say in that move. So if I see such expressions, it is easier for me to tell my meaning more effectively. And even the program can recognize my meaning then! (Student 63, interview transcript)

- Construction of new functional meaning

Says that for contrast he is writing "however, little is known about" because this should show a gap in previous research. (Student 54, observation transcript)

[changes: "A large number of scenarios can be gained to test forecasting tools with the scenario generation method." → "It is possible that if scenario generation method were adopted, a large number of scenarios could be gained to test forecasting tools." (Student 29, Camtasia transcript)

At the end, I changed more to make appropriate steps sentences. (Student 44, interview transcript)

4.6.3. Learner Fit

For the Learner Fit quality, the question was whether the IADE program and its automated feedback on discourse forms were appropriate for targeted students. Table 4.12 shows the types of data and analyses used to examine this aspect. A total of ten Yes/No, open-ended, and Likert-scale responses were considered. As it was mentioned in Section 4.3.4, one of the scale questions contained an item reflecting ten specific characteristics of IADE's feedback, which provided evidence of students' perceptions of the degree of helpfulness of the feedback they received.

Table 4.12. Aspects, data, and analyses of Learner Fit

Aspect of LF	*Data and analysis*	
Appropriateness of IADE and its feedback	*Quantitative*: yes/no responses; Likert-scale responses	comparison of response percentages
	Qualitative: open-ended survey responses; think-aloud protocols; observations; semi-structured interviews	manual analysis of responses; coding

Similar to the previous two CALL qualities analyzed, evaluation of Learner Fit also relied on introspective data from think-aloud protocols, semi-structured interviews, observations, and open-ended survey responses. In the coding scheme, Learner Fit was represented by the following categories:

- Facilitating problem-solving

  [moves the cursor over the same sentence again and again] I have no idea. Let me see other move 2 [goes to the annotated corpus] (Student 30, think-aloud protocol transcript)

  Reads, thinks, check the definitions, says that it was helpful and that it's clear now. (Student 32 observation transcript)

I think help options help me figure out stuff, like why color turn out different. (Student 61, interview transcript)

- Fit for individual learner characteristics

"I like that the feedback is not just something in general; it's meant for me personally and it reflect only my writing advantages and disadvantages. Plus, it tells me about the introductions in my discipline." (Student 58, interview transcript)

"I liked the part of the statistics. The numbers give me a clear view. I am so sensitive to numbers..." (Student 29, interview transcript)

"And I am sure I would revise and resubmit a lot of drafts because I am a perfectionist." (Student 44, interview transcript)

- Learner control

I know I'm right here. First, I introduce our research, then I talk about what we did, and then about our result. So, it can't be move 1. Maybe it looks like move 1? Why? Hm… maybe because sometimes author can talk about previous research and what other people did. Like, if I review a study from before, I can tell what they did and what they find. Yeah, I think so. But not here. I think my reader is smart enough to understand that I'm not talking about other study from somebody else here. I already said "Herein, we report our first study…", so it is clear enough. (Student 65, think-aloud protocol transcript)

- Appropriateness of feedback for task completion

"The feedback helps me to avoid overemphasizing some move and to improve." (Student 30, interview)

"I like that I can submit many times. It help me to see the real picture so I don't have to guess if I'm right or wrong." (Student 40, interview)

- Appropriateness of evaluation method

I like that the program doesn't give a grade or a score, because, if it's wrong for some sentences, the grade will be wrong, and then I be upset. But now it allows me to practice, but doesn't dictate what I should to do, or doesn't criticize me. It doesn't say – this is the last time and you get A, B, or C and so on. It just gives me the picture compared to other introductions in my field. (Student 30, interview transcript)

4.6.4. Impact

In order to find if the students had a positive learning experience with IADE or not, quantitative and qualitative data were analyzed again, as shown in Table 4.13. Percentages for 5 survey and 4 Likert-scale questions were calculated and compared.

Table 4.13. Aspects, data, and analysis of Impact

Aspects of I	*Data and analysis*	
Positive and/or negative influence	*Quantitative*: yes/no responses; Likert-scale responses *Qualitative*: open-ended survey responses; think-aloud protocols; observations; semi-structured interviews	comparison of response percentages manual analysis of responses; coding

Open-ended survey responses uncovered participants' perceptions of the types and consequences of Impact, and introspective insights were gained from coded think-aloud protocols, semi-structured interviews, observations, and. Following are the Impact categories used in the coding taxonomy:

- Positive learning experience

  I don't want to see any below or above average, so I feel like I have to make good changes. And I feel good to see good results (Student 61, interview transcript)

  He looks very satisfied when he sees the feedback because he sees improvement in both m1 and m2, which were problematic in the first draft. (Student 32, observation transcript)

- Negative learning experience

  I don't know what to do now. I tried so much, but nothing changed in the colors. (Student 40, think-aloud protocol transcript)

  You know, sometimes it's frustrating, especially when I change so many times, but the color is the same. (Student 43, interview transcript)

- Motivation

  Although the program's feedback shows improvement in move 2, (the % is close to average), he still decides to do more revision of that part because it is not as close to the average figure in his field as he would like it to be. (Student 32, observation transcript)

  He looks motivated and goes on revising. (Student 58, observation transcript)

- Cognitive involvement

  "the student seems to be deeply cognitively engaged", "the student does a lot of self-analysis", "the student is very concentrated", "the student looks motivated and goes on revising" (excerpts from observations transcripts)

  Mhm, ok. Hmm, yeah, let me think about it. Ok. (Student 40, think-aloud protocol transcript)

- Emotional involvement

  "awesome!", "yes, great!", "wow, finally!", "average, average, average, yes!", "no more orange!", etc. (excerpts from think-aloud protocol transcripts)

  She gets very excited every time the program analysis showed improvement both in the colored text and the numerical feedback. She even claps her hands!" (Student 28, observation transcript)

- Strategy development

  I think this program help me with every sentence. I can write a sentence and give it to the program to check if it is right move. If it is not right move, I can revise it. If it is correct move, I will write a new sentence and check with the program. So I can check every sentence and make sure that I make my structure and ideas clear. (Student 43, interview transcript)
  So, I just finished move 1. Now I will check move 2. I look at the result provided by the program, I change words and structure for each move, and I revise many times sentence by sentence for move 2. (Student 62, think-aloud protocol transcript)

- Influence on the revision process

  Usually, I revise my paper one time and that's it. It's because I don't know what else to revise. I make some changes and I think it's good. But this program helped me organize how to revise. (Student 54, interview transcript)

> *Yes, it's something new for me. I never think about function before. (Student 63, interview transcript)*

To sustain and elaborate on the themes that emerged from the qualitative data (e.g., participants' answers to open-ended survey questions, think-aloud/Camtasia transcripts, interview transcripts, observations transcripts, and first and last students' drafts), examples from the coded material will be provided in the results Chapters 5 and 6. To ensure the representativeness of sample examples, they were randomly extracted with SPSS, where a pre-defined number of cases was selected from a given coded subset of data. No random sampling was done to exemplify themes identified in small data subsets of up to 6 instances.

4.6.5. Scoring and reliability

4.6.5.1. Human rating and reliability

Two raters evaluated the quality of the rhetorical development of 16 first and 16 last drafts produced by the same participants. Before rating, they attended 18 hours of training held by the researcher, where they were acquainted with Swales' move schema and the class materials used in the English 101D course. The training covered not only explanatory and example materials, but also practice that consisted of annotating research article introductions as well as rating student introduction drafts, (other than the ones to be rated for research purposes). The trainees were first trained to annotate texts in order to provide the raters with a better understanding of the genre and the variation among disciplines. Then, they were trained to score student texts according to the criteria presented in Table 4.14.

Table 4.14. Rating scale

Rating criteria	*Effectively 3*	*Satisfactorily 2*	*Poorly 1*	*Score points*
1. Establishes a territory (stressing interest, indicating importance, and/or showing topic prominence; describing previous research; making topic generalizations – Move 1)				x
2. Establishes a niche (indicating a gap and/or a problem; raising a question; hypothesizing; presenting a justification – Move 2)				x
3. Occupies the niche (stating the purpose and/or describing present research, its methodology, and/or its results; stating its value; outlining its structure – Move 3)				x
Total score				x

Both aspects of the training included demonstration and explanation, individual work, and debriefing. Scores of 1, 2, and 3 were assigned to maintain similarity with the scores based on the program's automated evaluation. The texts were provided to the raters without any identification signs (except the student's discipline) or information that would disclose whether the drafts were first or last. To make better judgments about student drafts, the raters were also given the statistics for all the steps used in the three moves in each student's discipline. The statistics were helpful in that they informed the raters of the presence or absence of certain steps in the students' discipline as well as of the length of the steps. With this information, the raters were able to evaluate the students' moves more objectively.

Given that the raters each had to rate 32 drafts (16 first drafts and 16 last drafts) and to judge the quality of three moves in these texts, there were 96 overall cases for possible agreement, of which the raters agreed on 77 cases. In other words, the rater's agreed about 8

out of 10 times, their agreement being 80.21% (see Table 4.15). In cases of disagreement, the researcher acted as a third rater.

Table 4.15. Inter-rater agreement

	First draft		*Last draft*		*Total*	
	Agreement	Disagreement	Agreement	Disagreement	Agreement	Disagreement
Move 1	12 (12.5%)	4 (4.17%)	12 (12.5%)	4 (4.17%)	24 (25%)	8 (8.34%)
Move 2	13 (13.54%)	3 (3.12%)	13 (13.54%)	3 (3.12%)	26 (27.08%)	6 (6.24%)
Move 3	13 (13.54%)	3 (3.12%)	14 (14.58%)	2 (2.08%)	27 (28.12%)	5 (5.21%)
Total	38 (39.58%)	10 (10.42%)	39 (40.62%)	9 (9.37%)	77 (80.21%)	19 (19.79%)

A more robust measure of inter-rater reliability, Cohen's kappa coefficient (*k*), was calculated as well. Cohen's kappa was chosen because this measure uses the assumption that the same raters evaluated the same set of items, which was precisely the case in this sudy. The *k* coefficient was calculated as follows:

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}$$

where *P(A)* represents observed probability of agreement, and *P(E)* is the expected probability of agreement, i.e., chance agreement. Table 4.16 summarizes the numbers of scores that the raters agreed on for each move and the respective *k* coefficients. All the coefficients indicate a good level of agreement between the raters.

Table 4.16. Inter-rater reliability

		Move 1	*Move 2*	*Move 3*	*All moves*
Rater 1-Rater 2	N agreed	25	26	27	78
	k	.633	.719	.757	.714

4.6.5.2. IADE-human reliability

Final human scores on moves and the scores assigned based on IADE's automated analysis were used to measure IADE-human reliability. Of the same number of cases (96), IADE and human raters agreed on 70 cases, the agreement being 72.92% (see Table 4.17). IADE-human rater reliability was also estimated through calculations of the Cohen's kappa coefficient (*k*). Table 4.18 shows that this agreement was moderate for each and all moves.

Table 4.17. IADE-rater agreement

	First draft		*Last draft*		*Total*	
	Agreement	Disagreement	Agreement	Disagreement	Agreement	Disagreement
Move 1	11 (11.46%)	5 (5.21%)	13 (13.54%)	3 (3.12%)	24 (25%)	8 (8.33%)
Move 2	9 (9.37%)	7 (7.29%)	13 (13.54%)	3 (3.12%)	22 (22.91%)	10 (10.42%)
Move 3	11 (11.46%)	5 (5.21%)	13 (13.54%)	3 (3.12%)	24 (25%)	8 (8.33%)
Total	31 (32.29%)	17 (17.71%)	39 (40.62%)	9 (9.37%)	70 (72.92%)	26 (27.08%)

Table 4.18. IADE-human-rater reliability

Pair		*Move 1*	*Move 2*	*Move 3*	*All moves*
IADE-human raters	N agreed	24	22	24	70
	k	.582	.529	.607	.575

It is interesting to note that, unlike disagreement among human raters who almost equally disagreed on both first and last drafts, the disagreement between IADE and humans occurred mostly on first drafts (17.71% of 27.08% of cases of disagreement). However, this may not be entirely surprising because of the reliance of IADE's automated evaluation on n-grams; first drafts where such vocabulary was present might have scored higher. Human raters, on the other hand, focused more on how well the functional meaning was expressed, being more objective judges, and their scores were lower for first drafts in which the

vocabulary appropriate for certain moves was not very effectively employed. The disagreement rate between automated and human scores decreased for last drafts, where the participants must have made more appropriate use of both vocabulary and meaning. The next chapter in fact uncovers that this was indeed the case.

4.6.6. Chapter summary

This chapter outlined the mixed-methods research methodology employed. It first explained how the process-product direction (Warschauer & Ware, 2006) was applied to the context of this study in order to evaluate the effectiveness of IADE, which was introduced as an innovative AWE technology that supplemented a face-to-face course. It also explained how Chapelle's (2001) CALL evaluation framework overarched the conceptualization of the study. Then, the participnats were described in terms of academic background, age, gender, native language, English language proficiency, period of study at a US university, and research writing experience. Following the description of the data collection instruments (IADE's database, pre- and post-tests, survey questions, observations, and interviews) and of the data analysis tools (coding taxonomy used in Transana), the specifics of the procedure for the two data collection periods were detailed. Further, it was explained how data for the four CALL qualities investigated was analyzed and converged for the purpose of interpretation. Quantitative analyses included measures of learners' perceptions, of learning gains, of writing improvement, and of the role of practice and modified interaction for improvement. Qualitative data were coded, and evidence of emerging themes was triangulated. The chapter concluded with a section the reliability of human scores and on IADE-human reliability. The next two chapters introduce the results according to the research questions posed.

Chapter 5. Results on the Language Learning Potential of IADE

5.1. Introduction

This chapter presents the results concerning IADE's Language Learning Potential, which were obtained from multiple data sources and analyses described in Chapter 4. The findings are reported as they address the research questions about this CALL quality of the program:

- What evidence suggests that the feedback provided by IADE leads to students' noticing of and focus on discourse form?
- What evidence suggests that students learned the target discourse forms that were focused on during the interaction with IADE?

First, the results for the first question are introduced. They are based on both quantitative and qualitative data such as survey questions and frequency counts as well as computer screen recordings, think-aloud protocols, observations, and semi-structured interviews. The findings not only show that learners noticed and focused on the rhetorical moves in their discourse, but also provide an understanding of how this process unfolded and what it entailed. In other words, the results reveal that IADE's feedback instantly drew the participants' attention to the discourse form of their draft. Then, such focus made them notice negative evidence in their writing, motivating them to revise it. For that, they modified their output, experimenting with and testing different hypotheses by resubmitting their modifications for automated analysis and feedback, which then made them focus on discourse form again. Output modifications, especially unsuccessful ones, lead to more

deliberations and to modified interaction through IADE's Help Options, which enhanced learners' understanding of the moves and consequently led to more successful output modifications. The sequence and iterativity of this learning cycle became evident in introspective and observational data from 16 participants, and the scale and open-ended survey responses also pointed to the elements of this cycle. Therefore, it was concluded that this was a phenomenon characteristic of all participants' revision process.

Next, the results for the second research question are introduced through an examination of quantitative and qualitative evidence of improvement yielded by Likert-scale and survey questions as well as by computer screen recordings and automated and human scores on first and last drafts. Both perceptive data and draft scores showed that the participants improved the quality of their discourse by making output modifications at the level of content, vocabulary, grammar, structure, and mechanics. In addition, participants' learning gains are discussed based on the analysis of Likert-scale and survey questions and of pre/post-test scores. Specifically, the results here show that the learners believed that interaction with IADE and its feedback contributed to their learning of moves; t-tests confirmed these perceptions by finding statistical significance between the means of the scores on pre- and post tests. Finally, the roles of two variables - practice through multiple resubmissions of drafts and modified interaction through IADE's Help Options, are elaborated on. Qualitative data indicated that both practice and modified interaction were considered as very helpful by learners. Quantitative analyses revealed that practice was positively related to overall improvement of students' writing, while the relation between improvement and modified interaction was not statistically significant.

5.2. Noticing of and focus on discourse form

Evidence that learners noticed RA Introduction discourse forms and consequently directed their focus towards those while revising with IADE were obtained from 3 Likert-scale questions, 2 survey questions, and 16 think-aloud protocol/Camtasia transcripts, 16 observations, and 16 semi-structured interviews obtained from the same participants. Overall, these data sources yielded strong evidence of noticing of and focus on discourse form, which can be seen in Table 5.1 below.

Table 5.1. Overall evidence of noticing of and focus on discourse form

Data source	*N of participants*	*Data*	*Evidence*[2] *for LLP*	*No evidence for LLP*
Survey data: Likert-scale q-ns	88	Responses to:		
		Q-n 2: Did you focus on the moves?	92.05%	7.95%
		Q-n 6: Did you notice any words/expressions that seemed to be characteristic of certain moves?	100%	0%
		Q-n 7: Did the feedback provided by IADE help you see the weaknesses in your drafts?	90.91%	9.09%
Survey data: Yes/No & open-ended q-ns	83	Q-n 1: When working with IADE, has the program helped you focus on the moves? How? What exactly made you focus on them?	100%	0%
		Q-n 6: Has the program helped you notice anything about the moves/steps that you might not have paid much attention to before? How? What are these things?	89.16%	10.84%

[2]For survey questions, % indicates the percent of participants, in whose responses evidence of LLP was found.

Table 5.1. Overall evidence of noticing of and focus on discourse form (continued)

Data source	*N of participants*	*Data*	*Evidence*[3] *for LLP*	*No evidence for LLP*
		Transcripts of:		
Introspective data: Think-aloud protocols/Camtasia	16	16 Think-aloud/Camtasia recordings	484/1227 idea units	
Semi-structured interviews		16 interviews	63/233 idea units	
Observational data: Observations	16	16 observations	118/460 idea units	

5.2.1. Learner perceptions on noticing and focus on discourse form

In Likert-scale question 2, an overall 92.05% of 88 respondents indicated that they focused on form "somewhat" (47.7% or 42 students) or "a lot" (44.3% or 39 students). Only 6.80% (6 students) chose "a little" and 1.14% (1 student) chose "not at all". (See Figure 5.1)

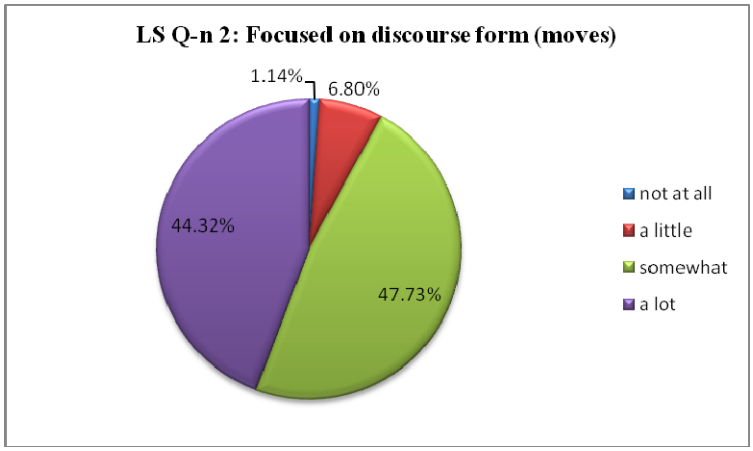


Figure 5.1. Participants' focus on discourse form in Likert-scale questionnaire

[3]For survey questions, % indicates the percent of participants, in whose responses evidence of LLP was found.

Answering survey question 1, 100% of 83 respondents also indicated that working with IADE helped them pay attention to and focus on the moves, providing a number of reasons for their judgment. As presented in Table 5.2, 98.8% of respondents mentioned that something triggered their focus on form.

Table 5.2. Results from survey question 1.

Triggers of focus on discourse form	*Participants*
Feedback	95.19%
- color-coded feedback	59.04%
- numerical feedback	12.05%
- color-coded and numerical feedback	24.1%
Definitions and examples	3.61%
Total	98.8%

The majority of respondents (95.19% or 79 students) explained that it was the feedback that made them focus on the moves; of which 59.04% (49 students) referred to the color-coded feedback, 12.05% (10 students) referred to the numerical feedback, and 24.1% (20 students) referred to both types of feedback, for example:

- Reference to color-coded feedback

  The colors help me focus on the moves, and know the distribution of the moves. (Student 57, survey, question 1)

  The different colors made me focus on the moves (Student 45, survey, question 1)

  Its colorful highlights brought my attention to my moves. (Student 54, survey, question 1)

  I could figure the sentences out with different colors. (Student 48, survey, question 1)

  The color is very striking for finding different moves. (Student 33, survey, question 1)

  The colors make me notice what move are written and what are not. (Student 65, survey, question 1)

I think it helped me focus on the moves by highlighting the different moves in colors. (Student 51, survey, question 1)

The colored sentences show me which moves the sentence belongs, is that what I want it to be? (Student 56, survey, question 1)

IADE *helped me to focus on different moves through the different colors that are used for each move. (Student 29, survey, question 1)*

The idea using colors was nice because different colors make me think about the structure of moves. (Student 27, survey, question 1)

This program helped me focus on the move a lot. Because it is very easy to use and can tell which one is move 1, 2, 3 very easily. (Student 24, survey, question 1)

Color helped me to distinguish my moves. (Student 31, survey, question 1)

It visualizes in color what I need to add on or reinforce - so yes. (Student 34, survey, question 1)

- Reference to numerical feedback

It helps me focus on the move with the statistics number and the comparison with average in my field. It makes me focus on the percentage of move in the introduction. (Student 85, survey, question 1)

When the program tells me that certain move in my article is below the average in my discipline, I start to focus on that move. (Student 59, survey, question 1)

How? It could determine whether I am focusing in one move or another depending on the percentages. (Student 30, survey, question 1)

The program has helped me a lot. It focused me on the moves and how they are commonly distributed my discipline. (Student 8, survey, question 1)

Percentages of moves made me focus on them. (Student 22, survey, question 1)

It helped a lot in the weights and distribution of the moves. (Student 66, survey, question 1)

It gives me idea on what is the average length of the move, so I pay more attention to what is not enough. (Student 74, survey, question 1)

- Reference to color-coded and numerical feedback

The color feedback tells me which types of moves I have and also I know whether the move is in average of my discipline (Student 50, survey, question 1)

The color and the % helped me. (Student 42, survey, question 1)

> *Yes, the program helps me focus on the moves. The* IADE *help me with the colors and the percent that I need to follow. I focused in the percent of the moves. (Student 47, survey, question 1)*
>
> *The program helped me focus on the moves by providing the colors and how much I was away from the average on my discipline in each move. (Student 31, survey, question 1)*
>
> *Yes it was very helpful to focus, when we are able to look at the moves and % and analyze its distribution. (Student 39, survey, question 1)*
>
> *The feedback is all about the moves by colors and % so that it makes me concentrate on moves which I was not concerned a lot. (Student 7, survey, question 1)*
>
> *Colors and statistics (Student 70, survey, question 1)*

The other 3.61% (3 students) mentioned that accessing the definitions and examples of moves provided in IADE's Help Options helped them better understand and focus on their own discourse form. The remaining 1.2% (1 student) made only general comments about IADE's helpfulness, writing *"it helped me in understanding the structure of introduction and how to balance the moves"* (Student 28, survey, question 1).

Likert-scale question 7 was a question directly related to IADE's feedback, and it revealed that the majority of the participants thought that the feedback was facilitative in that it helped them notice weaknesses in the rhetorical development of their drafts. Answers to the four given options were distributed as follows: 40.91% (36 students) chose "a lot", 50% (44 students) chose "somewhat", 7.95% chose "a little", and 1.14% (1 student) chose "not at all" (see Figure 5.2).

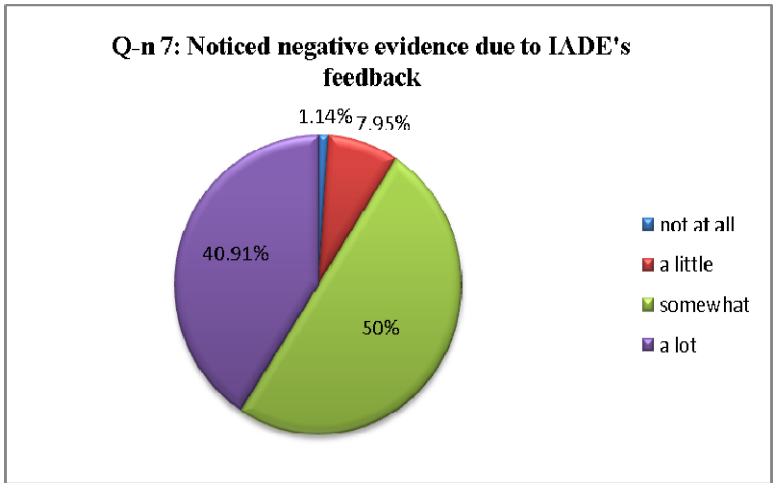


Figure 5.2. Results from Likert-scale question 7

Furthermore, in response to survey question 6, 89.16% (74 participants) pointed that IADE also helped them notice certain peculiarities of the rhetorical moves in research article introductions that they had not paid attention to during their corpus explorations in class and while writing their first draft, some pointing to focus on automated feedback. Table 5.3 presents the ideas mentioned by the participants.

Table 5.3. Results from survey question 6

Aspects noticed	*Participants*
Move-specific phraseology	57.83%
Negative evidence	12.05%
Learning of moves	7.23%
Feedback	12.05%
Total	89.16%

More specifically, 57.83% (48 students) mentioned noticing that certain move-specific phraseology typically signals the function of a step/move very clearly and that a given function of a step within a move can be fulfilled in a number of different ways through different vocabulary choices. It is not surprising to see such answers here considering that the majority of participants indicated in Likert-scale question 6 that they noticed vocabulary

indicative of certain moves. More precisely, 38.64% noticed this “a lot”, 52.27% “somewhat”, 9.09% “a little”, and nobody chose to say “not at all”. Following are some examples of survey question 6 open-ended responses containing evidence about participants’ noticing of move-specific words/expressions.

> *… this program makes me notice the expression of the sentences. (Student 57, survey, question 6)*
>
> *Yes, the key words indicating the moves. (Student 53, survey, question 6)*
>
> *… after that I noticed that there are some key words. (Student 37, survey, question 6)*
>
> *In specific I would say, in picking appropriate words. (Student 70, survey, question 6)*
>
> *Yes, principally specific words that could mean one move/step or another. (Student 8, survey, question 6)*
>
> *Yes, the words I should write in each move. (Student 55, survey, question 6)*
>
> *I realized that there are specific words or expressions researchers use for specific moves or steps. I actually notice that many researchers say the same function with different key words. (Student 41, survey, question 6)*
>
> *When I didn't know what's wrong with my move 2, I looked at many examples and I saw that I can express my gap with different expressions. So I learned more vocabulary for move 2. (Student 31, survey, question 6)*
>
> *The pop messages in the corpus of* IADE *helped me see that I can use different ways for my centrality, but in fact say the same thing. (Student 5, survey, question 6)*

Participants’ answers to this question also contained evidence for other claims about Language Learning Potential. For example, 12.05% (10 students) stated that they noticed negative evidence in their output, e.g., missing moves, a difference in their move distribution compared to the patterns in their discipline, or a mismatch between the intended function and the actual verbal realization of it.

> *It makes me analyze my own sentence and see what are the problems of each sentence. (Student 38, survey, question 6)*

This way can help people pay attention to how to distribute reasonably. (Student 40, survey, question 6)

Different length, style in my academic subject, the balance based on jornals in each field. (Student 80, survey, question 6)

I notice the construction of introduction using this program. Before that, I mainly wrote the present condition, problems, and purpose as memory serves. I could see that there are many types and steps in moves to logically write research through inherent characteristics of this program. (Student 34, survey, question 6)

For example, centrality in move 1. I know the sentence is move 1, but feedback say it move 2. So I think and I make sure which step the sentence is, then I can modify. (Student 44, survey, question 6)

From the feedback of the program, I found some sentences were written as move 3, for example. But the program notified that it is a sentence for move 2. (Student 37, survey, question 6)

It helped me notice some steps that were not paid much attention. For example, some sentences are thought to be in Move 2, but are analyzed to be Move 1. Then I know I should revise them and make them look like sentences in Move 1. (Student 51, survey, question 6)

Also, 7.23% (6 participants) expressed confidence about their newly acquired knowledge, writing that they learned the moves better after revising with IADE.

I didn't know these moves well before I actually revised my draft with the program. Now, I know. (student 52, survey, question 6)

I didn't recognize the differences between Move 1 and Move 3 very well, but now I have learned that Move 1 show the information on the topic will focus on whereas Move 3 need to pay more attention on the author's research. (student 30, survey, question 6)

I didn't quite understand these moves before. But now, I was able to see how I write them so I know them. (student 35, survey, question 6)

The program helps me see things I usually don't really think about, and that made me learn the complicated moves. (student 6, survey, question 6)

The responses of the other 10 students (12.05%) were statements that indicated focus on feedback (4 on numerical feedback, 4 on color-coded feedback, and 2 on both types of

feedback). The remaining 10.84% of participants' responses did not contain relevant evidence - 2 students (2.41%) were not sure whether they noticed anything in addition to what the feedback pointed to, and 7 students' answers (8.43%) were irrelevant to this question.

5.2.2. Introspective data on noticing and focus on discourse form

In addition, the combined think-aloud protocol and Camtasia transcripts recorded for 16 participants provided evidence that all these participants referred to the form of their discourse in one way or another while thinking aloud. Of all the idea units identified in these transcripts (1,227), 484 (39.45%) idea units were coded for language learning potential, and 260 (53.72%) of those were coded for focus on form, with an average of 16.25 idea units per student. Figure 5.3 shows the number of focus on form idea units per student.

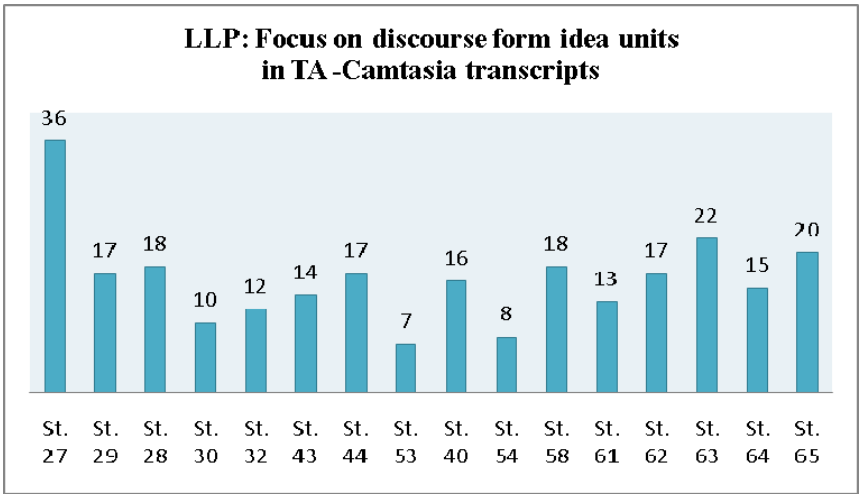


Figure 5.3. Focus on discourse form in think-aloud protocol and Camtasia transcripts (per student)

These coded data confirmed that the color-coded and numerical feedback generated by IADE was indeed the driving factor that made the students focus on the form of their discourse as it

was noted by many of them in their answers to survey question 6. Consider the following examples:

> *First, it's good to know that I have all three moves. (Student 27, think-aloud protocol and Camtasia transcript)*
>
> *I see it's, m1 and m2 it's basically in the front, and most of m3 is in the end of the introduction. And m1 and m2 are... they are intertwined, and m3 are basically at the end. (Student 29, think-aloud protocol and Camtasia transcript)*
>
> *… so it should be put here, put before the sentences which describe the homogeneous advantage of spin coating. (Student 28, think-aloud protocol and Camtasia transcript)*
>
> *so, right now the structure at move level seems to be fine (Student 30, think-aloud protocol and Camtasia transcript)*
>
> *So, my moves are like this… I didn't expect this part would be included in m1 [mouse over blue text]. (Student 32, think-aloud protocol and Camtasia transcript)*
>
> *It shows that there is lots of move 1. (Student 43, think-aloud protocol and Camtasia transcript)*
>
> *[Opens AC, scrolls up and down looking at the overall distribution of moves/colors, then looks at his own colored text.] (Student 44, think-aloud protocol and Camtasia transcript)*
>
> *Well, these are m1 and m2. m3 is just one sentence.(Student 53, think-aloud protocol and Camtasia transcript)*
>
> *[goes up to the color-coded feedback, highlights blue sentences in between green ones] (Student 40, think-aloud protocol and Camtasia transcript)*
>
> *um, m2 is a little... is scattered inside there in m1. (Student 54, think-aloud protocol and Camtasia transcript)*
>
> *I think there is... there are only two moves in my introduction. (Student 58, think-aloud protocol and Camtasia transcript)*
>
> *[looks at the color-coded feedback] A lot of blue. (Student 61, think-aloud protocol and Camtasia transcript)*
>
> *[turns back to the color-coded text. Slowly scrolls down.] (Student 62, think-aloud protocol and Camtasia transcript)*
>
> *First of all, let me check the statistics with my introduction. [cursor over numerical feedback] ok. (Student 63, think-aloud protocol and Camtasia transcript)*

I have several blue sentences after the m3, after the sentence which is indicated as m3, and then again m3. (Student 64, think-aloud protocol and Camtasia transcript)

In addition, these transcripts also provided clear insights into the positive consequences of students' focus on discourse form, which were mentioned by the participants in their scale responses. As a result of such focus, the learners noticed negative evidence in their draft, which was also confirmed by their answers to Likert-scale question 7, and consequently modified their written output. Since their modifications were not always successful upon first attempt, the learners consulted IADE's Help Options and compared the materials they accessed with their own draft, and that led to a better understanding of the moves and the steps within the moves as well as to improved changes in the draft. Figure 5.4a shows that, apart from focus on form, 24.17% (117) idea units pointed to noticing of negative evidence, 14.05% (68) idea units - to output modification, and 8.06% (39) - to enhanced understanding of moves. Figure 5.4b presents the number of idea units per student.

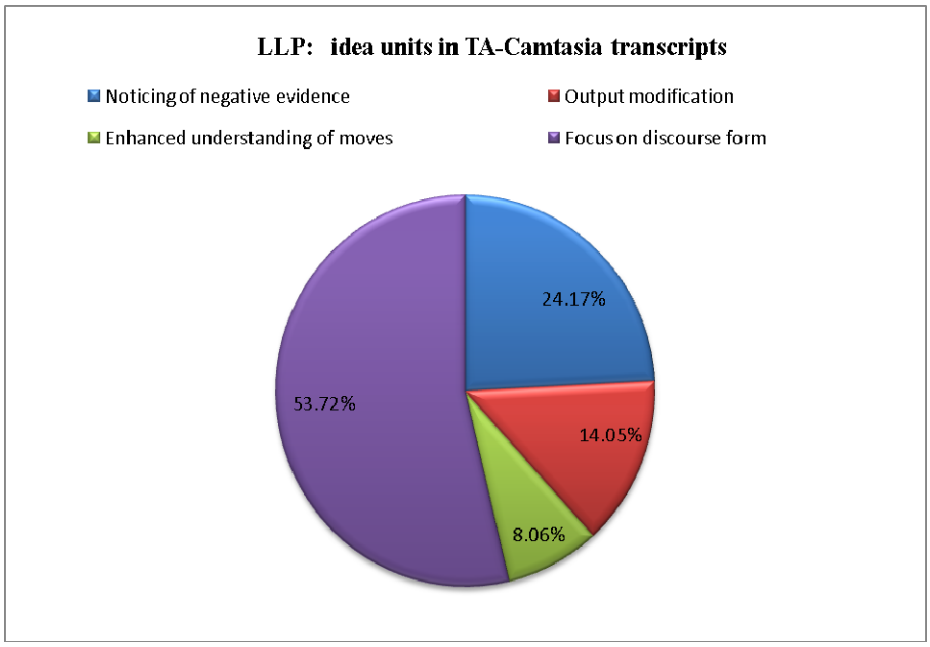


Figure 5.4a. Idea units coded for LLP in think-aloud protocol and Camtasia transcripts

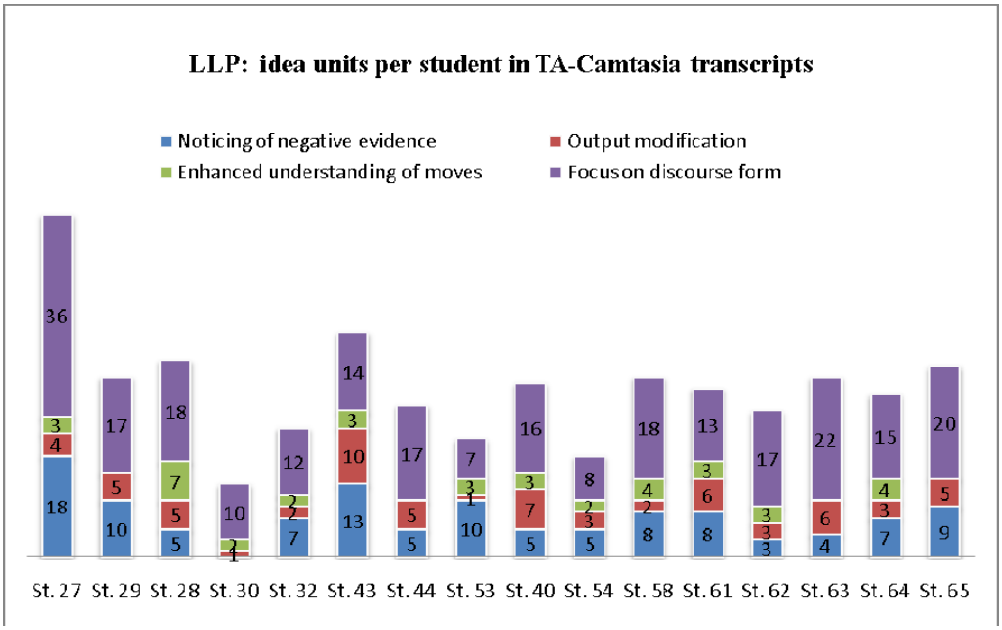


Figure 5.4b. Idea units coded for LLP in think-aloud protocol and Camtasia transcripts (per student)

Here are some quotes from the think-aloud and Camtasia transcripts that exemplify instances when the students noticed negative evidence, modified their drafts, and seemed to gain a better understanding of the rhetorical moves.

- Noticing of negative evidence

  Hmm... There is a question that for m1, my previous research was identified as m3 ... since, indeed, I have my comment on it. (Student 27, think-aloud protocol and Camtasia transcript)

  I was trying to say, to state the value of the present research, but it turned out to be m2. (Student 29, think-aloud protocol and Camtasia transcript)

  Now I'm looking at my draft. The first sentence [highlights it] um,I think um it is the background of the problem, so it should be m2. (Student 28, think-aloud protocol and Camtasia transcript)

  so, "but" is in uh, oral English, not very academic. (Student 30, think-aloud protocol and Camtasia transcript)

  I thought um the… previous research condition is limited, so it doesn't ... this sentence should be m2, but from this program it's m1. (Student 32, think-aloud protocol and Camtasia transcript)

… stating the value of present research. I think I missed that part. And this sentence is duplicating the previous one. (Student 43, think-aloud protocol and Camtasia transcript)

…but it still recognizes not like a hypothesis but like a ... maybe another objective. (Student 44, think-aloud protocol and Camtasia transcript)

I think in my introduction the moves ... m1 and m2 are a little... it's not that centralized. (Student 53, think-aloud protocol and Camtasia transcript)

So this part I want it to be m3, but uh it still displays in m1. I don't know why. See here, this part. [highlights a blue part of the text] in my mind I wanted to write a method. So I think it describes the method, but it's m1. (Student 40, think-aloud protocol and Camtasia transcript)

Um, m2 is a little... is scattered inside there in m1. Not so good. (Student 54, think-aloud protocol and Camtasia transcript)

I see there is... there are only two moves in my introduction. (Student 58, think-aloud protocol and Camtasia transcript)

… and it is still below average in my discipline. Uh, I need to add something else because in m3 I only outline the structure, but nothing else. (Student 61, think-aloud protocol and Camtasia transcript)

And I was trying to continue the the problem … this sentence here, but... it is marked like m1 because uh... (Student 62, think-aloud protocol and Camtasia transcript)

I believe I should change something here because it's not about the statistical issues, statistical problems, so it was just more, more information... it's kind of random. (Student 63, think-aloud protocol and Camtasia transcript)

For this sentence… [reads a green sentence] should be famous problems in our test case. I don't know which move it belongs to. It seems that it's just a general statement. And I don't know why it is classified as m3. (Student 64, think-aloud protocol and Camtasia transcript)

I cannot make sure this is m1. It's blue, and I think I need move 3, so I want to check the definitions. (Student 65, think-aloud protocol and Camtasia transcript)

- Output modification:

Just now I just added some sentences in the m3. First of all I added a summary of my paper. Secondly, I added a brief review of my structure and then, finally, I added

something about the results. (Student 27, think-aloud protocol and Camtasia transcript)

I changed it to "The results will give the industry highlights on how to integrate UCD with HCI interface designing better." (Student 29, think-aloud protocol and Camtasia transcript)

… so I'll put it here, put before the sentences which describe the homogeneous advantage of spin coating. (Student 28, think-aloud protocol and Camtasia transcript)

[adds: "Previous research tried to separate some amino acids with various conditions, but there is no information about the separation phenomena without acidic modifiers."] If I add this sentence, I have to change ... [in "Although additives can improve the selectivity of chiral drugs, it is hard to separate these additives after separating D and L-form." deletes "Although".] (Student 32, think-aloud protocol and Camtasia transcript)

[changes: "The contribution of MetNetGE to metabolic network visualization is:" --> The contributions of our proposed system to metabolic network visualization are:"] (Student 43, think-aloud protocol and Camtasia transcript)

[scrolls up in the revision box, fixes some articles, makes up for the cohesion, and adds: "Mesocarp is composed by tryglicerols (TG), carotenes and vitamins; TG, represent between 90 -95% of the total composition of lipids (Sundram et al., 2003), carotenes and, vitamin E during the developing process of the fruit."] (Student 44, think-aloud protocol and Camtasia transcript)

[begins making changes: "He Xu believes that" --> "Scholars in China believe that". (Student 53, think-aloud protocol and Camtasia transcript)

If I use in "this model" because the sentence before is something about review, so I should change this into a modal. [makes the change] Yes. And this sentence here. I can combine this sentence into a long one. [changes: "In our model, tungsten cones with different sizes were used to enhance the optical field right under which a silicon substrate with 1 um thickness and 1 nm away from the tip was set." submits] (Student 40, think-aloud protocol and Camtasia transcript)

[changes the problematic red sentence to begin with "the importance of". deletes a part of the sentence. Modifies more: "In fact, the importance of cement fineness has been elevated to the heat of hydration.". (Student 54, think-aloud protocol and Camtasia transcript)

[adds: "The definition of RAFT is..."]. (Student 58, think-aloud protocol and Camtasia transcript)

Structure, yeah, this is the structure. I can move it right before the last sentence. [cuts the sentence and pastes it later in the text.] (Student 61, think-aloud protocol and Camtasia transcript)

[adds: " Although there are no man antecedents in Iowa, nowadays we have a framework for the search of new alternative solutions for that."] so, I added a sentence uh that uh mention that ummm... because... about the lack of antecedents uh of this kind of approach, so ... I will submit again. (Student 62, think-aloud protocol and Camtasia transcript)

I'm going to include another reference I know by heart. [adds: "Rosen (1993) also introduced a variance estimator for ordered frames."] (Student 63, think-aloud protocol and Camtasia transcript)

[changes: " A class in the model is declared to be a different type asyn." --> "A class in our model is declared to be a different type asyn."] (Student 64, think-aloud protocol and Camtasia transcript)

[highlights the previously modified sentence. adds "it means" to "It means the absorption mechanism is different from activated carbon." thinks. reads this sentence again, cursor over. changes: "It means that, compared to activated carbon, the absorption mechanism is different from activated carbon."] (Student 65, think-aloud protocol and Camtasia transcript)

- Enhanced understanding of moves:

I realized that … in all the examples, in the m3, there's no previous research. (Student 27, think-aloud protocol and Camtasia transcript)

The third sentence is ... [highlights and thinks] I think it's a way to solve this problem. So, it may be m3. I don't remember, I have to check. [opens the HO definitions] Now I'm seeing the steps in m3. No, it's not m3. Hmmm... It says 'studies of smth have been hampered by the fact that this exhibits weak activity in ...", so it displays the drawbacks of the study. hmmm... Oh, in my sentence I only said it's a big problem, but there's no drawbacks in my discipline. (Student 28, think-aloud protocol and Camtasia transcript)

So, I use m2 to uh... highlighting a problem or indicating a gap? [checks definitions] Yeah, this is m2, step 2, highlighting a problem. (Student 30, think-aloud protocol and Camtasia transcript)

I said I want to use the question rising for m2, but I found from the statistical results that there is no question rising from the corpus. And I found that indicating a gap is much more a factor than other m2 steps in my corpus. (Student 32, think-aloud protocol and Camtasia transcript)

[reads from the HO example and emphasizes "A key antecedent of public .... need to be better integrated in our understanding".] hmmm.... [reads another example and emphasizes: " understanding of the .... is crucial". highlights is crucial to determine"] ok, I have some ideas. (Student 43, think-aloud protocol and Camtasia transcript)

[goes up to the color-coded feedback. reads, cursor over. highlights a red sentence. goes down to the revision box] well, is this sentence the... [highlights the sentence] it's just showing that the previous research uh ignores something, some important topic. so it should be... that could be m3. [goes to the HO definitions] or it's still m2. [thinks] indicating a gap. Yeah, that's what is. (Student 53, think-aloud protocol and Camtasia transcript)

I don't remember much m2. [opens HO definitions] I think to find some examples. [opens the AC in the HO. clicks on red sentences] oh, I know! In m2 if I want... there is a step... indicating a gap. In this step describes some problem in the previous research. Oh, I know! It's natural. It should be follow the review. So, I don't think... I... doesn't matter where this step will be located. (Student 40, think-aloud protocol and Camtasia transcript)

This one also, this one m1. This is... this is uh... umm... I think this is uhh... [explains what happens in that phenomenon, cursor over a red sentence] this is very important, then I think this one maybe m1, centralize. Ok, I understand. (Student 54, think-aloud protocol and Camtasia transcript)

Ok, so it's the importance of the knowledge... ok, stressing interest... indicating the importance... so... [reads the examples provided there] so I think I'm [indistinct word] about claiming centrality and um, the uh the step is 3. (Student 58, think-aloud protocol and Camtasia transcript)

Let me see some examples. Yeah. [opens the AC. Reads m3, cursor over green text] we identify... [indistinct, reads in quiet voice, cursor over] oh! opens another annotated text, finds green text, highlights parts of it] present tense... yeah... [highlights "is developed". thinks. reads in half voice. Highlights "was developed". Sighs deeply. Clicks on green sentences in AC to see the steps. looks at the definitions] summarizing method... announcing principal outcomes... it's developed first... so, this is the method. Then structure, yeah, this is the structure. (Student 61, think-aloud protocol and Camtasia transcript)

[opens AC and looks at and reads carefully examples of m2] hmm... so I guess I should focus more on the problem that we find, and uh... why we should develop this research. So I should add more about the why we should uh... make this research and this development project... in order to get more statement of the problem. (Student 62, think-aloud protocol and Camtasia transcript)

> *I don't think it meet any of the uh... categories. They are not. Just a general statement about a tool we are going to use for the test. Whether our project set is right property or not... nothing to do with project... I don't know... [opens definitions. thinks] it seems that it related to m3. (Student 64, think-aloud protocol and Camtasia transcript)*

A close analysis of the students' thoughts and actions indicates that IADE's feedback seemed to have stimulated a cycle (Figure 5.5) that they appeared to follow during revision once they received the automated feedback on their drafts. The head of the cycle was learners' focus on discourse form; one cycle ended with successful output modification of a certain discourse element reflected in the feedback upon resubmission, and a new one began with learners' focus on another part of their text. Inside this cycle, there appeared to be another integrated mini-cycle, during which the learners noticed negative evidence in their work products and made multiple attempts to understand its nature and to make corrective changes.

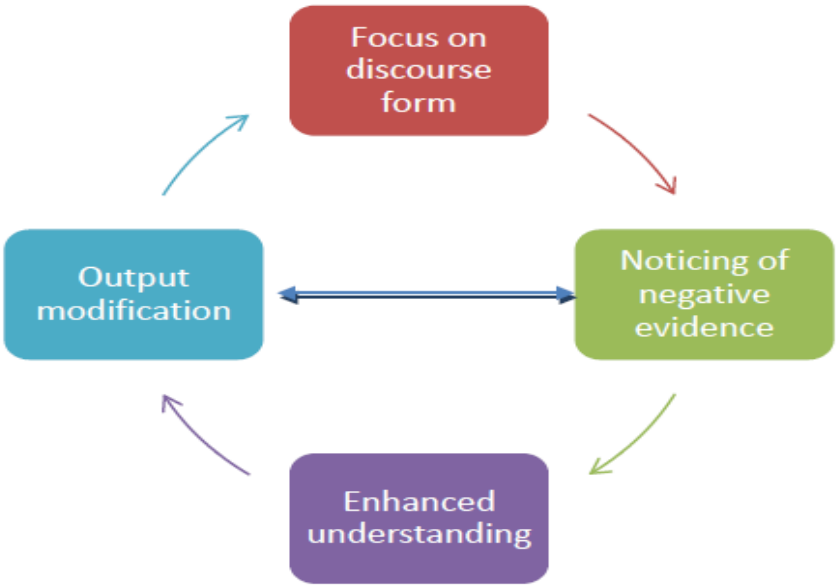


Figure 5.5. Revision cycle stimulated by IADE's feedback

Figure 5.6 presents an excerpt from the think-aloud/Camtasia transcript for Student 27, which shows how the student goes through the stages of this cycle. First, the feedback, both color-coded and numerical, prompts his focus to the distribution of the three moves in his RA introduction. Then, the student notices negative evidence clearly pointed to by the numerical feedback, and then, while reading the color-coded text, realizes that a function of his introduction was identified by IADE differently than intended. Having noticed this negative evidence, the student makes a change in his text based on a personal hypothesis, and, upon re-submission, sees that his hypothesis was faulty. This, thus, motivated him to think more and consult IADE's Help Options, which lead to an enhanced understanding, in this case, of the discourse norms in his particular discipline. With a better understanding of that, the student modified his output again – this time successfully, which the student was able to see by focusing on form again. His revision process continued with another iteration of the cycle.

Focus on discourse form:
The program tell me that the blue is m1, red is m2, and green is m3. First, it's good to know that I have all three moves. Let's see what's in each move. Discipline... 66% belong to m1... above average. Yeah, so, that means in m1 I have about 20% above the average, but below the maximum.

Noticing of negative evidence:
[looks at the numerical feedback] So there is more room for improvement. [reads the feedback prompt - try to revise this move] OK, now I know what I'm going to do because I'm above the average for m1, I'm below, substantially below the average for m3. So, maybe now I need to shorten m1 and elaborate m3. [reads one of his sentences. highlights part of a sentence] Hm… my previous research was identified as m3. Why it happened like this? If it appears at the beginning of this paragraph… so I guess maybe the program just sees it as m1 according to its location. Let's see if I change the position, what will happen.

Figure 5.6. Example of the cycle stimulated by IADE's feedback (think-aloud protocol and Camtasia transcript, Student 27)

Output modification:
[goes down to the revision box. reads his text again. Copy-pastes a piece of text to a place at the beginning]. OK, I already changed it, changed the location of it.

Noticing of negative evidence:
Hmm... Although I changed the location, it is still regarded as m3. All right, now I know that it is not because of the location. Maybe it's because of the language…

Enhanced understanding:
Let's see some examples. [goes to the help options, Annotated Corpus (AC), and looks at one annotated introduction] [whistles quietly] M3 in this example does not have the previous research review, so I will look at other examples. [opens other annotated texts trying to find review in green] I realized that all the examples, in the m3, there's no previous research. So, I guess, that's a problem… I can just keep the previous research in m1 and try to add something in m3, like … let's see… methods… add some parts like the structure or the summary of the outcomes. So…

Output modification:
Now I'm focused on the last paragraph of my introduction. I'm trying to add something. [reads the last part of his text. Goes to IADE's colored text. Goes back to the revision box and adds: "For its methodology, this study used content analysis to find out the historical flow of Chicago downtown. The paper is structured in accordance with the five stages of Florida's creative economy. The results show that, indeed, the Chicago downtown, although has a relatively short history, matches Florida's theory."] Just now I just added some sentences in m3. First of all, I added a summary of my paper. Secondly, I added a brief review of my structure, and then, finally, I added something about the results. Now let's see how the program will do my next draft. [submits]

Focus on discourse form:
Now the program told me that it indeed detected some additional m3, and yeah!.. I'll try to revise more, I think. [reads on. highlights a sentence and reads carefully. Highlights it again]

Noticing of negative evidence:
And here I find another misunderstanding. [reads the sentence] My intent is to ... I guess it could be m2 or m3, but not m1, so I'm thinking, if it is m2 or m3. [reads the sentence aloud again and takes time to think] It is hypothesis, yeah.

Output modification:
I'm trying to use the word "assume" as an indicator of m2 hypothesis. [Changes: "So it is possible to set downtown Chicago into Florida's five economic transitions" --> "So this paper assumes that it is possible ..." Let me see if it works and if the program can recognize this change.

Figure 5.6. Example of the cycle stimulated by IADE's feedback (think-aloud protocol and Camtasia transcript, Student 27) (continued)

Noticing of negative evidence:
Hmm... no? it is still recognized as m1. Let's see some examples. I'll find out what kind of indicators words the previous papers. The annotated corpus...

Enhanced understanding:
[opens texts in the AC] This example text does not have this part [meaning step], so I'll look at another one. [clicks on red sentences, most of which are 'problem'. Since he doesn't find hypothesis, he opens another annotated text] I'm looking for some indicator words for hypothesis in m2. I already went through two example texts, but both of them m2 just have justification and statement of problems. I didn't find indicator words for hypothesis, so I'll just keep looking for it. No. gap, gap, justification, problem, gap ... I went through several other examples, but still I didn't get indicator words for hypothesis in m2. So, I'm wondering why the previous research papers do not have that. Maybe they just don't include hypothesis in my field? [opens the stats in Help Options and confirms his idea] I guess so… because only 0.14% for hypothesis. So, according to my observation, maybe it's true that the other authors in my field don't have hypothesis included in m2. Then I will do more what they do. Let me look at my results.

Output modification:
I'm looking at the last sentence ... and to see if I can make some changes and the program can respond to these changes. Maybe I can try to be specific and say that "the results of this paper". Yeah, let's try to do that. [makes the change" "The results show that..." --> "The results of this study show that..." and submits]

Focus on discourse form:
Yep! All right! Now it changed to m3, it's what I mean. So that's good! Ok, so now what I'll do… I'll analyze sentence by sentence and ... and then I can change when things wrong.

Figure 5.6. Example of the cycle stimulated by IADE's feedback (think-aloud protocol and Camtasia transcript, Student 27) (continued)

Next, from a total of 460 idea units in the observations data, 118 (25.65%) were coded for language learning potential, with an average of 7.37 idea units per participant. This data showed that all 16 observed participants exhibited behavior indicative of language learning potential and also contained evidence of the pattern described above (see Figures 5.7a and 5.7b).

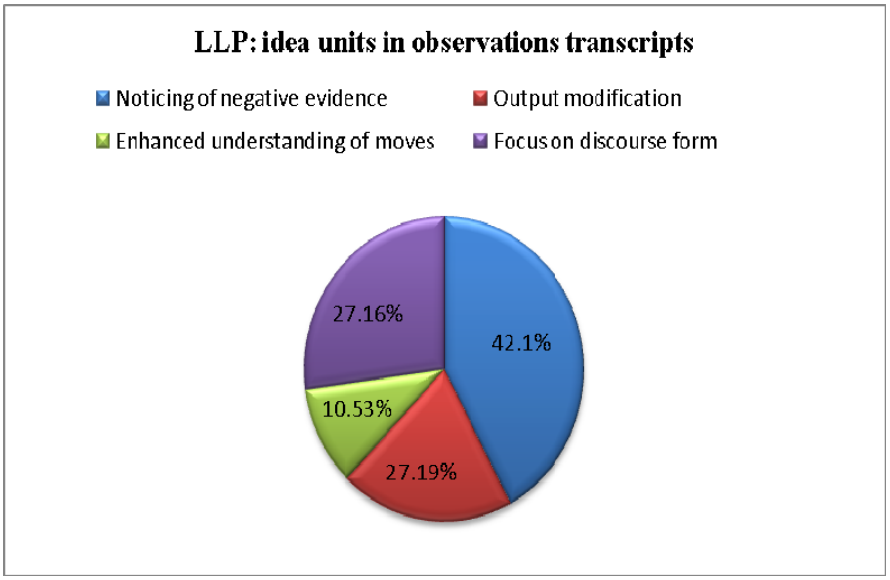


Figure 5.7a. Idea units coded for LLP in observations transcripts

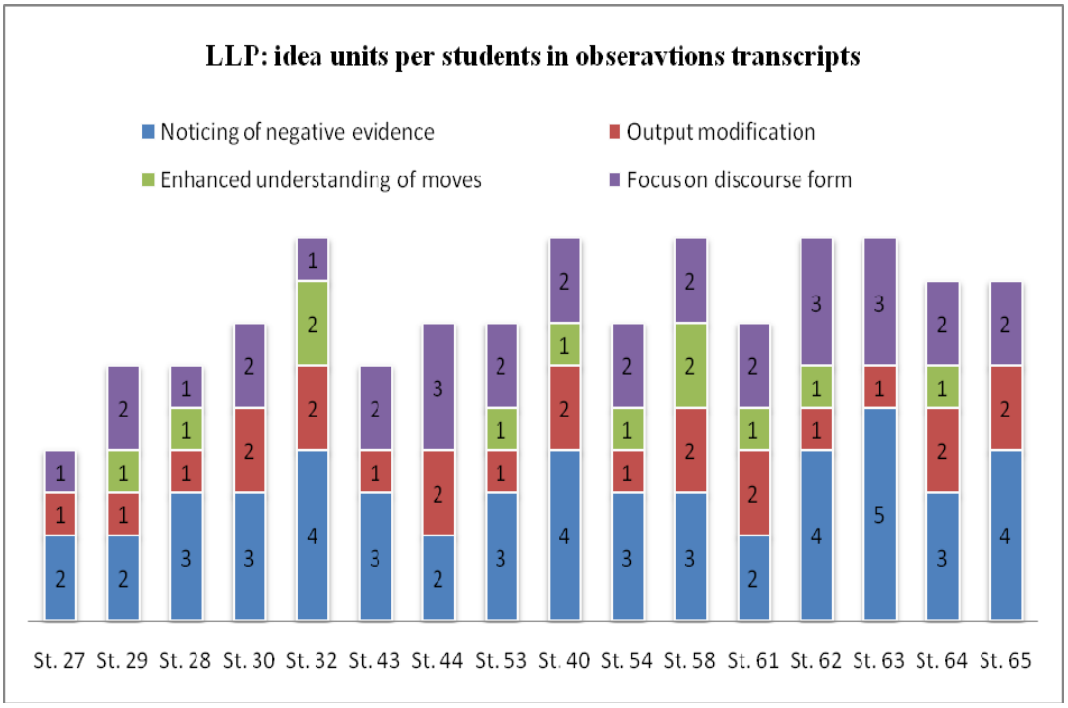


Figure 5.7b. Idea units coded for LLP in observations transcripts (per student)

Also, consider the example in Figure 5.8 where an excerpt of the researcher's observation notes for Student 28 were coded for focus on discourse form, noticing of negative evidence,

enhanced understanding, and output modification, which appeared to occur in the same order as depicted in Figure 5.5.

Focus on discourse form:
Looks carefully at the color-coded text returned by IADE. Frowns.

Noticing of negative evidence:
The first thing she says upon receiving IADE's feedback is “Wow! No move 2 and lots of m1!” She immediately says that it's not what she means.

Output modification:
Decides that she definitely needs to revise. Moves closer to the computer and takes a position for focused work. Makes changes to her text in the revision box.

Noticing of negative evidence:
Goes back to the color-coded feedback. Highlights parts of the color-coded text. Notices that she actually meant m2.

Enhanced understanding:
Thinks. Begins reading and thinking about the first sentence. Accesses the AC (annotated corpus) for examples of m2 Problem. After seeing examples, thinks that she may need to elaborate on the Problem. But also thinks that it may be better not to start with m2 because she didn't see it at the beginning of any annotated introductions. The second sentence is also meant as m2, but appears as m1. She thinks that because it's too long it doesn't reflect the problem clearly.

Output modification:
Decides to split the sentence into 2 --> one as Generalization and the second as Problem. Makes changes in the two sentences. Submits.

Focus on discourse form:
Is happy to see a successful change. Now m2 appears as she wants it. Reads and thinks.

Noticing of negative evidence/ Output modification:
Says that some information is not very essential, so she may want to delete it. Then, wants to combine the rest of the sentence, part of which she has just deleted, with the next sentence because, as she explains, that one shows the advantage and is therefore more important.

Enhanced understanding:
Classifies her next sentence as move/step; then checks with the definition of m3 and reads the examples there. She now has plenty of ideas on how to manipulate with her information and content (many times she says “maybe I can do this; maybe I can do that...”).

Figure 5.8. Example of the cycle stimulated by IADE’s feedback (observations transcript, Student 28)

Output modification:
Decides to combine sentences. Then deletes the detailed sentences about the reference. Modifies the placement of sentences for more coherence of ideas. Also deletes something that seems redundant to her.

Focus on discourse form:
Looks carefully at colors, then at numbers…

Figure 5.8. Example of the cycle stimulated by IADE's feedback (observations transcript, Student 28) (continued)

The same cyclical pattern became prominent in the retrospective semi-structured interview data where the researcher's questions probed for an understanding of participants' actions. This data consisted of 233 idea units produced by the same 16 participants, of which 27.04% (63 idea units) were related to the learning cycle in question. For example:

Focus on discourse form and **noticing of negative evidence:**
Researcher (following up on the observation note [doesn't like the distribution of her moves]): What did you mean when you said that your introduction is "not that centralized"?
Student 43: I mean not coherent, not logical, not as should be.
Researcher: How do you know?
Student 43: Because I saw how colors are all over the place.

Output modification:
Researcher (following up on the observation note [decides to try combining a green and a blue sentence]): Why did you decide to combine those two blue and green sentences at the beginning?
Student 43: Oh, because it wasn't right. They should both be move 1, but I didn't write like that. I thought if I combine it helps, but it doesn't.
Researcher (following up on the answer and on the observation note [combining doesn't work, so he checks examples in the AC and makes a number of changes at lexical level]): So that's why you changed that part several times?

Enhanced understanding:
Student 43: Yes. I look at the corpus and I didn't see "we" in move 1, but I saw "we" in move 3. So I finally understood the problem there.

Figure 5.9. Example of the cycle stimulated by IADE's feedback (interview transcript, Student 43)

Researcher (following up on the answer and on the observation note [repeatedly looks at examples of moves/steps in HO and then highlights/reads his own sentences]): Yes, I noticed that you used the Help Options to see examples and then you went back to your own sentences. Why were you doing that?
Student 43: I just compare corpus with my sentence. Because I can see that they are different, and then I think why it's different and I find out why. I like also definitions because if I go to definitions and examples, then I understand little things … like why I can' use "we" in move 1. I understand that "we" tells about this research, not research in general, right? I didn't pay attention before.
Researcher: So that helps?

Output modification:
Student 43: Yeah, sure. Then I know finally what to change, and it works.

Focus on discourse form:
Researcher: How do you know?
Student 43: Because I see the colors and I know what to do next to make it good.

Figure 5.9. Example of the cycle stimulated by IADE's feedback (interview transcript, Student 43) (continued)

Not only did the interview data help contain the elements of the same cycle, but, when coded, the questions and answers also seemed to point to the same sequence of this cycle. What is particularly interesting is that participants' answers to questions about their actions during one stage of the cycle provided evidence about the effect of the preceding stage. This can be seen in the raw example provided in Figure 5.9 above, and to make it more clear, relevant questions and answers are synthesized and simplified below:

Question: How did you know that there were imperfections in your introduction? → **Noticing of negative evidence**

Answer: Because I saw the distribution of my moves in color and in percentages. → **Focus on discourse form**

Question: Why do you think you better understood how to build each move more effectively? → **Enhanced understanding**
Answer: Because I saw what and where my problems were and tried to clarify them. → **Noticing of negative evidence**

Question: Why did you make changes to particular parts of your introduction? → **Output modification**
Answer: Because I could see my problems and because I could better understand the purpose, function, and realization of each move. → **Noticing of negative evidence** + **Enhanced understanding**

Question: Why did you pay attention to your moves after resubmission? → **Focus on discourse form**
Answer: Because I wanted to see how good they were. → **Output modification**

Figure 5.10 summarizes these generalized ideas, showing how the participants explained their thoughts and actions in retrospect:

- noticed negative evidence because of focus on discourse form;
- acquired a better understanding of discourse conventions because of noticing of negative evidence
- modified output because of noticing of negative evidence and acquiring a better understanding of discourse conventions
- focused on discourse form because of wanting to verify the quality of the modified output and to continue revision of other parts of the text.

Considering this perceptive data obtained in interviews, it can be inferred that the relation between the four elements of the identified cycle is sequential and causative.

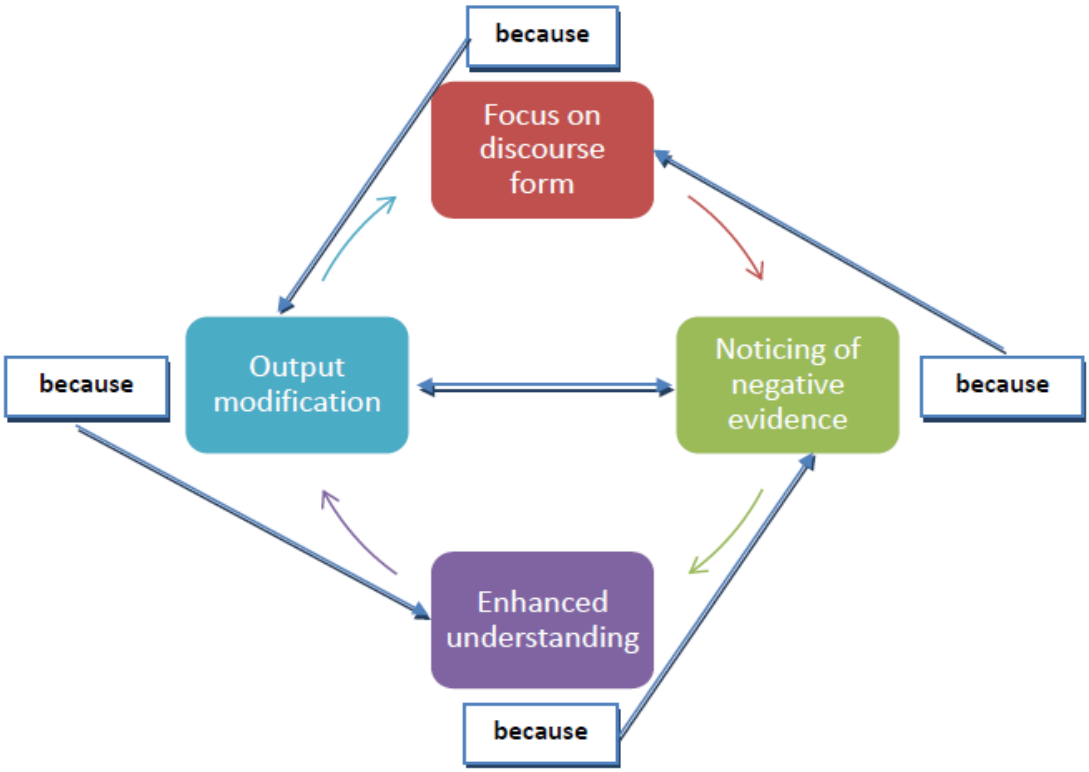


Figure 5.10. Sequential and causative relation between the elements of the revision cycle

5.3. Improvement in rhetorical development

Evidence suggesting that the participants improved the rhetorical development of their research article introduction discourse as a result of their interaction with IADE and its automated feedback found support in several data sources listed in Table 5.4. More specifically, such evidence was identified through analyses of students' perceptions expressed in their responses to two Likert-scale and survey questions, of automated evaluation of 210 first and last drafts, of human rating of 32 first and last drafts, and of the output modifications manually identified in 16 think-aloud/Camtasia transcripts and in 32 first and last drafts (see Table 5.4).

Table 5.4. Overall evidence of improvement

Data source	*Nr of participants*	*Data*	*Evidence for LLP*	*No evidence for LLP*
Survey data: Likert-scale q-ns	88	Responses to: Q-n 5: Did your RA Introduction writing skills improve?	100%	0%
Survey data: Yes/No & open-ended q-ns	83	Q-n 8: Do you think you improved your skill of writing a research article introduction? Why?	92.8%	7.2%
Scores: IADE's automated evaluation of 1st and last drafts	105	Drafts: 210 first and last drafts	t-test Move 1: p =.000 Move 2: p =.000 Move 3: p =.000 Length: p =.000	
Human scores of 1st and last drafts	16	32 first and last drafts	Move 1: p =.000 Move 2: p =.000 Move 3: p =.000	
First and last drafts	16	32 first and last drafts	285 output modifications	
Introspective data: Think-aloud protocols/Camtasia	16	Transcripts of: 16 Think-aloud/ Camtasia recordings	77 output modifications	

Participants' answers to the Likert-scale and survey questions provided insights about how they perceived their own improvement, IADE's automated evaluation and human scores measured the improvement in participants' work products, and the analysis of participants'

output modifications captured with Camtasia and identified in 16 participants' first and last drafts added evidence of improvement in the rhetorical quality of the students' discourse.

5.3.1. Learner perceptions on improvement

Responding to Likert-scale question 5, all the participants reported that they believed they improved their skill of writing a research article Introduction section to a certain degree: 26% (23 students) thought they improved "a lot", 55% (48 students) thought they "somewhat" improved, and 19% (17 students) thought they improved "a little". None of the participants perceived no improvement at all (see Figure 5.11).

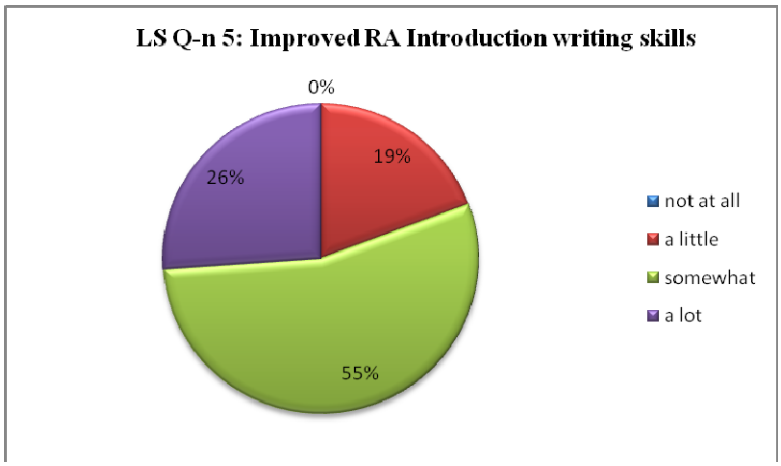


Figure 5.11. Participants' perceptions of improvement in Likert-scale questionnaire

In survey question 8, 92.8% (77 participants) of 83 respondents answered "yes" when asked if they thought they improved their Introduction writing skills, and other 7.2% (6 participants) were uncertain, saying:

> *I don't know, I have never written such paper before. (Student 38, survey question 8)*
> *I am not very sure about that (Student 19, survey question 8)*
>
> *I don't know yet (Student 17, survey question 8)*
>
> *Not quite sure how much for now (Student 5, survey question 8)*

I wish I know exactly. (Student 80, survey question 8)

Maybe. This program just tells me the moves, not how much I improve. Also I don't think I spent enough time. I think if I spend more time, I'll improve more. (Student 74, survey question 8)

The explanations of 10.39% (8 participants) of the 77 participants' who believed they improved were generic and not indicative of particular reasons, while the other 89.61% of participants' responses regarding why they thought their Introduction writing skills improved formed two major themes. They appeared to judge improvement (1) based on what they thought caused it and (2) based on perceived or observed outcomes. Among the causes of improvement, 48.05% (37 participants) mentioned:

- IADE's ability to direct their attention to the discourse form of their draft (29.87%, 23 participants)

  I improved my writhing skill by using the program and focusing the moves (Student 57, survey question 8)

  I put more focus on Moves. (Student 42, survey question 8)

  I saw my introduction in more organized form (Student 7, survey question 8)

  I think because I noticed more and more on the structure of introduction. (Student 37, survey question 8)

  I pay more attention to the introduction, and the moves. The program instruct me to see and write the paper. (Student 12, survey question 8)

- Their focus on move-specific vocabulary during revision (6.49%, 5 participants)

  I paid more attention on logical expressions. (Student 61, survey question 8)

  The keys words I saw many I think improved my paper. (Student 52, survey question 8)

  It thought me some words to use in my writing. (Student 22, survey question 8)

I found academic words and logical construction of a sentence. (Student 34, survey question 8)

I tried and checked to control my words and sentences (Student 63, survey question 8)

- Opportunity for practice through multiple resubmission (11.69%, 9 participants)

  More practice was helpful to improve. (Student 58, survey question 8)

  I practiced a lot with IADE*, so I improved. (Student 85, survey question 8)*

  If I practiced more, I might improve more. (Student 41, survey question 8)

  I wrote and submitted my research introduction for several days until now, so I guess my skill got improved. (Student 34, survey question 8)

The outcomes that 41.56% (32 participants) believed were indicative of improvement in their skills were:

- Enhanced understanding and knowledge of Introduction sections' rhetorical conventions (25.97%, 20 participants)

  I learned the steps and moves in the introduction section and I can use them in my own writing. (Student 52, survey question 8)

  I think I improved my skill of writing a little, because I learned how to organize the sentences for different moves. (Student 51, survey question 8)

  I know I improved because I have a better idea about how to organize the introduction now. (Student 77, survey question 8)

  I had not had no idea about how to write research article introduction before, but now I know it (Student 70, survey question 8)

  I understand the moves so I am good at and comfortable writing introductions. (Student 33, survey question 8)

- Better quality of end products (15.58%, 12 participants)

  I see the differences in quality of writing introduction and other parts of research article. (Student 55, survey question 8)

My ideas became more clear and logical when I finished. (Student 47, survey question 8)

I could write more logically in my last time I submitted. (Student 73, survey question 8)

I think so, because my major professor said my introduction is really good. And also I have now a different look and I can know what is good and what is not good even when I read others papers. (Student 31, survey question 8)

These themes in participants' perceptions of why improvement occurred are shown in Figure 5.12. They resonate with the learning cycle identified in the think-aloud, observation, and interview data. Namely, focus on form and enhanced understanding are exactly the same elements as in that cycle, and practice through multiple resubmissions implies output modifications, which being eventually successful result in better quality of end products.

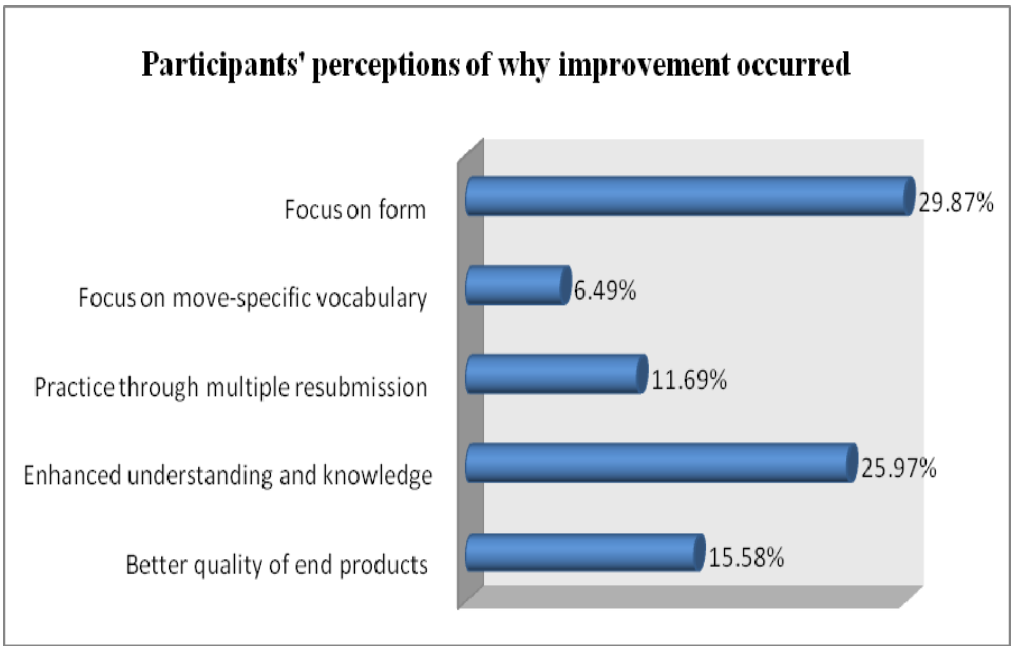


Figure 5.12. Participants' perceptions of why improvement occurred

5.3.2. Improvement in rhetorical development from first to last draft based on IADE's automated evaluation

IADE automatically evaluated students' drafts through comments for each move, which were generated based on percentages comparing students' texts to the corpus in their discipline. The comment that reflected desired performance is "This is about average in your discipline"; the other comments indicated how far the students' drafts were from targeted rhetorical norms of their particular discipline. IADE's evaluation of moves and draft length is given in Table 5.5, showing that the level of desired performance increased from first to last draft: for move 1 from 41% to 94.3%, for move 2 from 34.3% to 90.5%, for move 3 from 15.2% to 86.7%, and for length from 57.1% to 81%.

Table 5.5. IADE's automated evaluation of moves and length of first and last drafts

IADE's evaluation	*First draft*	*Last draft*
Move 1		
Far below average	8.6% (9 drafts)	0% (0 drafts)
Below average	6.6% (7 drafts)	0% (0 drafts)
About average	**41% (43 drafts)**	**94.3% (99 drafts)**
Above average	33.3% (35 drafts)	5.7% (6 drafts)
Far above average	10.5% (11 drafts)	10.5% (11 drafts)
Move 2		
Far below average	19.1% (20 drafts)	0% (0 drafts)
Below average	21.9% (23 drafts)	2.9% (3 drafts)
About average	**34.3% (36 drafts)**	**90.5% (95 drafts)**
Above average	13.3% (13 drafts)	6.6% (7 drafts)
Far above average	11.4% (12 drafts)	0% (0 drafts)
Move 3		
Far below average	23.8% (25 drafts)	0% (0 drafts)
Below average	45.7% (48 drafts)	9.5% (10 drafts)
About average	**15.2% (16 drafts)**	**86.7% (91 drafts)**
Above average	6.6% (7 drafts)	3.8% (4 drafts)
Far above average	8.6% (9 drafts)	0% (0 drafts)
Length of draft		
Far below average	4.8% (5 drafts)	0% (0 drafts)
Below average	31.4% (33 drafts)	13.3% (14 drafts)
About average	**57.1% (60 drafts)**	**81% (85 drafts)**
Above average	3.8% (4 drafts)	1.9% (2 drafts)
Far above average	2.9% (3 drafts)	3.8% (4 drafts)

IADE's automated evaluative comments were then transformed into scores as described in Section 4.6.1.2 of Chapter 4. Specifically, IADE's comment 'about average' was considered equal to a score of 3, 'below average' or 'above average' – to a score of 2, and 'way below average' or 'way above average' – to a score of 1. The scores assigned to each move and the length of students' first and last drafts are provided in Table 5.6, which shows that the distribution of all these elements appears to be better in last drafts compared to first drafts since considerably more last drafts were evaluated with the highest score of 3 (move 1 – 97.1%, move 2 – 92.4%, move 3 – 87.6%, and length – 83.8%). Overall, of 420 total possible scores, the highest score of 3 was assigned 159 times (37.9%) to first drafts and 379 times (90.2%) to last drafts. Similarly, considerably fewer scores of 1 and 2 were assigned to last drafts compared to first drafts.

Table 5.6. Scores for first and last drafts based on IADE's automated evaluation

Element	*Score 1*		*Score 2*		*Score 3*	
	#First draft	#Last draft	#First draft	#Last draft	#First draft	#Last draft
Move 1	15.2% (16)	0% (0)	41% (43)	2.9% (3)	43.8% (46)	97.1% (102)
Move 2	29.5% (31)	0% (0)	34.3% (36)	7.6% (8)	36.2% (38)	92.4% (97)
Move 3	33.3% (35)	0% (0)	52.4% (55)	12.4% (13)	14.3% (15)	87.6% (92)
Length	7.6% (8)	0.9% (1)	35.2% (37)	15.2% (16)	57.1% (60)	83.8% (88)
Total	21.4% (90)	0.2% (1)	40.7% (171)	9.5% (40)	37.9% (159)	90.2% (379)

In view of these results, one may be tempted to conclude that the students' work products improved, particularly at move level. Additional analyses were carried out to support this assumption. The descriptive statistics for first and last drafts' automated evaluation converted to scores for Move 1, Move 2, Move 3, and draft length are given in Table 5.7. It is worth mentioning that a positive sign is the fact that the standard deviations in each case decreased on the last drafts compared to the first drafts, which means that some

students, who did not do so well on their first draft, performed better after revising with IADE. T-tests comparing these means demonstrate that improvement from first to last draft was statistically significant with a probability value of .000 for all the moves and the length (see Table 5.8).

Table 5.7. Descriptive statistics for improvement (based on IADE's evaluation)

		Mean	*N*	*Std. Deviation*	*Std. Error Mean*
Pair 1	Move1_last_draft	2.97	105	.167	.016
	Move1_first_draft	2.29	105	.717	.070
Pair 2	Move2_last_draft	2.92	105	.267	.026
	Move2_first_draft	2.04	105	.808	.079
Pair 3	Move3_last_draft	2.88	105	.331	.032
	Move3_first_draft	1.81	105	.666	.065
Pair 4	Length_last_draft	2.83	105	.403	.039
	Length_first_draft	2.46	105	.665	.065

Table 5.8. T-tests results for moves and length (based on IADE's evaluation)

		Paired Differences					*t*	*df*	*Sig. (2-tailed)*
		Mean	*Std. Dev.*	*Std. Error Mean*	*95% Confidence Interval of the Difference*				
					Lower	*Upper*			
Pair 1	Move1_first_draft - Move1_last_draft	.686	.711	.069	.548	.823	9.88	104	.000
Pair 2	Move2_first_draft - Move2_last_draft	.886	.800	.078	.731	1.041	11.34	104	.000
Pair 3	Move3_first_draft - Move3_last_draft	1.067	.750	.073	.921	1.212	14.57	104	.000
Pair 4	Length_last_draft - Length_first_draft	.371	.609	.059	.254	.489	6.25	104	.000

For a better understanding of the degree of rhetorical improvement, the scores assigned to the moves and the length were used to classify improvement into four categories:

- Considerable improvement (improvement from first to last draft from a score of 1 to a score of 3)
- Noticeable improvement (improvement from first to last draft from scores 1,2 to scores 2,3 respectively)
- No improvement (no change of lower to higher scores from first to last draft)
- No need for improvement (score of 3 for both first and last drafts)

Table 5.9 and Figure 5.13 both detail the degree of improvement from first to last draft, listing the number of drafts and percentages for each category of improvement per each move and length. The 'noticeable improvement' and 'no need for improvement' categories were the most prominent ones. On average, 39.5% of students improved their drafts noticeably and 18.5% considerably. Most 'noticeable' and 'considerable' improvements were made at the level of Move 3 (52.4% and 28.6%). Move 3 may have been easier for students to develop since, to accomplish the functions of this move, they had to describe their own work, which is something they knew very well. Length was the aspect that was least improved; however, 55.2% appeared to need no improvement. The percentages of students whose first draft discourse elements saw no improvement were the lowest (0.9%, 3.8%, 7.6%, and 14.3%). The "no need for improvement" category had a relatively high percentage, and there may be two possible explanations for this result. It may have been the case that some participants' drafts were good to start with, or, it may be a limitation of IADE's automated analysis. IADE evaluates the rhetorical quality of discourse based on the unigrams, bigrams, and trigrams that have the highest probability of occurrence, and therefore the scores might have been higher when such vocabulary was present in the text.

Table 5.9. Degrees of improvement in moves and length (based on IADE's evaluation)

Improvement category	*Move 1*	*Move 2*	*Move 3*	*Length*	*Average improvement*
Considerable improvement	13.3% (14)	25.7% (27)	28.6% (30)	6.6% (7)	18.5% (19.5)
Noticeable improvement	43.8% (46)	38.1% (40)	52.4% (55)	23.8% (25)	39.5% (41.5)
No improvement	0.9% (1)	3.8% (4)	7.6% (8)	14.3% (15)	6.7% (7)
No need for improvement	41.9% (44)	32.4% (34)	11.4% (12)	55.2% (58)	35.2% (37)

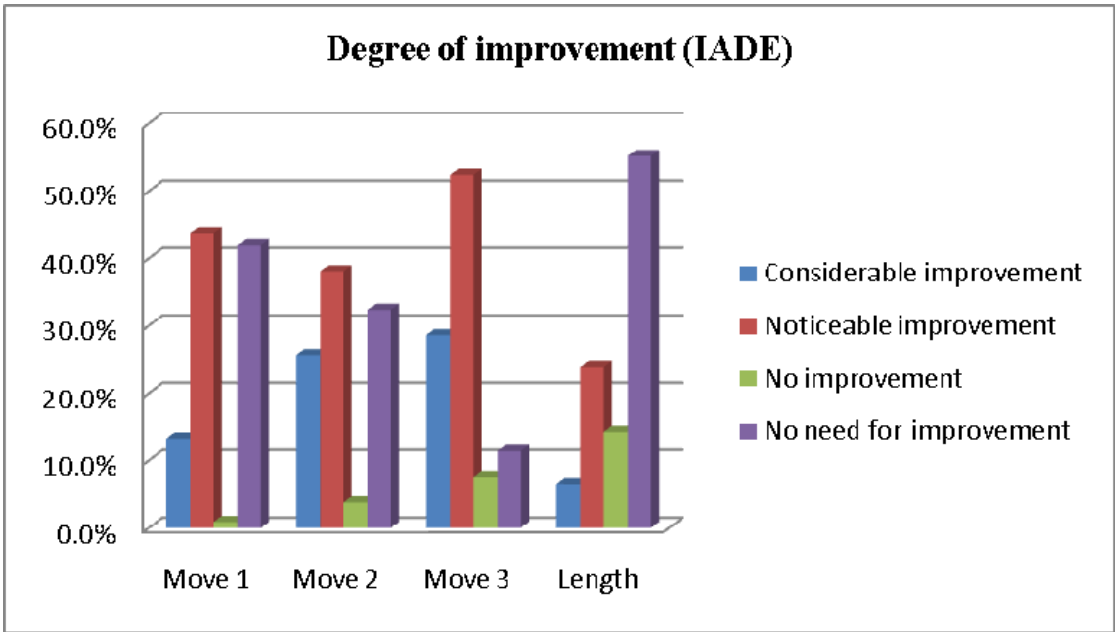


Figure 5.13. Degrees of improvement in moves and length (based on IADE's evaluation)

5.3.3. Improvement in rhetorical development from first to last draft based on human rating

The human scores for the 32 first and last drafts reveal that first drafts were weaker in rhetorical quality and that they became stronger after revision. As shown in Table 5.10, a total of 47.9% of moves received a score of 1 in first drafts and 0% in last drafts. The number of 2 scores also decreased from 45.8% in first drafts to 25% in last drafts. The number of the highest scores of 3 increased by 12 times (from 6.3% to 75%).

Table 5.10. Scores for 32 first and last drafts (based on human rating of 32 first and last drafts)

Element	*Score 1*		*Score 2*		*Score 3*	
	#First draft	*#Last draft*	*#First draft*	*#Last draft*	*#First draft*	*#Last draft*
Move 1	4 (25%)	0 (0%)	12 (75%)	3 (18.7%)	0 (%)	13 (81.3%)
Move 2	11 (68.8%)	0 (0%)	5 (31.2%)	5 (31.2%)	0 (%)	11 (68.8%)
Move 3	8 (50%)	0 (0%)	5 (31.2%)	4 (25%)	3 (18.8%)	12 (75%)
Total	23 (47.9%)	0 (0%)	22 (45.8%)	12 (25%)	3 (6.3%)	36 (75%)

This encouraging piece of evidence for improvement as judged by human raters is supported by the descriptive statistics and the t-test results comparing the first and last scores for each move (see Tables 5.11 and 5.12).

Table 5.11. Descriptive statistics for improvement (based on human rating of 16 first and last drafts)

		Mean	*N*	*Std. Deviation*	*Std. Error Mean*
Pair 1	Move1_last_draft	2.81	16	.403	.101
	Move1_first_draft	1.75	16	.447	.112
Pair 2	Move2_last_draft	2.69	16	.479	.120
	Move2_first_draft	1.31	16	.479	.120
Pair 3	Move3_last_draft	2.75	16	.447	.112
	Move3_first_draft	1.69	16	.793	.198

Table 5.12. T-tests results for moves (based on human rating of 16 first and last drafts)

		Paired Differences					t	df	*Sig. (2-tailed)*
		Mean	*Std. Dev.*	*Std. Error Mean*	*95% Confidence Interval of the Difference*				
					Lower	*Upper*			
Pair 1	Move1_last_draft - Move1_first_draft	1.063	.574	.143	.757	1.368	7.408	15	.000
Pair 2	Move2_last_draft - Move2_first_draft	1.375	.619	.155	1.045	1.705	8.883	15	.000
Pair 3	Move3_last_draft - Move3_first_draft	1.063	.772	.193	.651	1.474	5.506	15	.000

The means for all three moves increased in last drafts, and the difference between those means is significant with a *p* value of .000, which is the same as the one obtained when comparing the scores based on IADE's automated analysis.

Keeping the parallel with the analysis of scores based on IADE's evaluation, the rhetorical improvement here was also classified into 'considerable', 'noticeable', 'no need', and 'no improvement'. As it can be seen in Table 5.13 and Figure 5.14, the highest average percentage belongs to 'noticeable' improvement (54.38%) and the lowest to 'no need for improvement' (6.25%), a finding which is similar to the degrees of improvement that resulted from the automated evaluation.

Table 5.13. Degrees of improvement in moves and length (based on human rating of 32 first and last drafts)

Improvement category	*Move 1*	*Move 2*	*Move 3*	*Average improvement*
Considerable improvement	3 (18.75%)	7 (43.75%)	5 (31.25%)	5 (31.25%)
Noticeable improvement	11 (68.75%)	8 (50%)	7 (43.75%)	8.7 (54.38%)
No improvement	2 (12.5%)	1 (6.25%)	1 (6.25%)	1.3 (8.12%)
No need for improvement	0 (0%)	0 (0%)	3 (18.75%)	1 (6.25%)

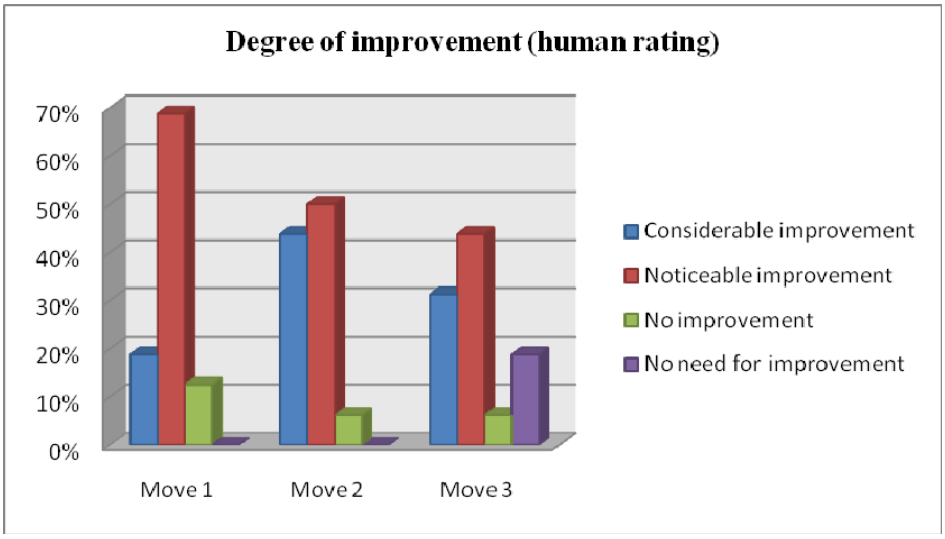


Figure 5.14. Degrees of improvement in moves and length (based on human rating of 16 first and last drafts)

The second high percentage of 31.25% belongs to ‘considerable improvement’ and the second lowest to ‘no improvement’ (8.12%). These results are somewhat opposite to the numbers for the same improvement categories based on IADE’s evaluation, but this outcome could have resulted from the environment in which the participants interacted with the program. More specifically, improvement categories based on human scores reflect the performance of those participants whose work was observed and recorded, which might have influenced them to invest more efforts into the revision process. Automated evaluation, on the other hand, captured the work of all the students, who may not have necessarily addressed everything they could have. In addition, although IADE’s analysis was modeled on human annotations, compared to human rating IADE-human agreement was somewhat weaker (see Chapter 4, Section 4.6.6).

5.3.4. Improvement in rhetorical development from first to last draft based on Camtasia transcripts

Improvement in students’ writing performance was also reflected in the modifications the 16 students made during revision, which were captured in Camtasia screen recordings and also manually identified by the researcher in a direct comparison of those participants’ first and last drafts. Camtasia transcripts contained 66 idea units (17%) coded as output modification, and these units contained overall 77 changes. Manual analysis of 32 students’ drafts revealed that 31 sentences remained unchanged (13%), and 200 sentences (87%) were modified and contained 285 changes from first to last draft. As uncovered in the analysis, students made changes at different levels: content, lexical, grammatical, structural, and mechanics. The frequency of these output modifications are presented in Table 5.14.

Table 5.14. Output modifications in Camtasia and first to last draft data

Level of output modification				*Total*	
		Camtasia	*1st-last drafts*	*Camtasia*	*1st-last drafts*
Content	Additions	15	66	18 (23%)	80 (28%)
	Deletions	2	3		
	Modified ideas	1	11		
Lexical	Move-specific	28	62	38 (49%)	121 (42%)
	Non-move-specific	10	59		
Grammar	Verb tense/form	4	12	5 (7%)	14 (5%)
	SV agreement	1	1		
	Plurals		1		
Structure	Sentence	7	50	10 (13%)	60 (21%)
	Paragraph	3	10		
Mechanics	Citation format	6	6	6 (8%)	10 (4%)
	Punctuation		4		
Total output modifications				77 (100%)	285 (100%)

- Content-level modifications

The modifications were considered to be done at the level of content when new material was added to the draft or when old material was either deleted or modified with more accurate information. In Camtasia data, 18 modifications (23%) were made in content, and 80 instances (28%) of content modifications were found in 16 first to last drafts. Additions were the most frequent content modifications identified in these two data sources – 15 in the former, and 66 in the latter. Deletions constituted 2 and 3 instances, respectively. The remaining number of content changes were typically modifications of ideas, which were not sufficiently tackled or not entirely precise (1in Camtasia data and 11 in first to last drafts). For example, Student 28 modified the content by expanding the idea of one sentence and providing more specific information about old and new approaches:

> *Not like old approaches which treat user cognitive process as predictable, UCD makes user the center during the entire design process. → There are several*

> *methodologies to implement different systems. Early ones tend to treat users' cognitive processes as predictable and quantifiable (Xu, 1996; Prett, 1998; Gros, 1999). New ones like User-Centered Design focus on constant conversations with users (Sholt, 2008). Users are made the center of designing process - this ensures that what is designed is what users need. (Student 28, first to last draft modification)*

Random examples of additions and deletions are presented below:

> *I'm adding m1 step 2. [Adds: "With easier" Deletes. Adds "The design of HCI interface is normally" Deletes "is normally" Adds "is normally focused on how to make computers understand human orders better. With easy" Replaces "orders" with "instructions". Thinks. Contunues: "With easy operating systems or graphical interface, users will find it" replaces "easy" with "user-friendly". Continues: "users will find it easy to give orders. So the computers can" Adds "precise orders". Continues: "can react to the users better". OK, I added some sentences here. (Student 29, Camtasia transcript)*
>
> ***Added →** Pressure sensor is one of the most successful applications of MEMS technology. Capacitive pressure sensor is known to have no turn-on temperature drift, high sensitivity and robust structure, and is less sensitive to side stress and other environmental effects. (Student 61, first to last draft modification)*
>
> ***Deleted →**Especially, the selectivity of tryptophan mixture is big enough to separate two enantiomers efficiently. (Student 32, first to last draft modification)*
>
> ***Deleted →** Research about the relationship between China's FDI development and the environmental protection varies among scholars in other countries. (Student 53, first to last draft modification)*

Content modifications appeared to always be move-specific. For instance, in the example below, Student 54 changed the content of the sentence so that it reflects a function of Move 1 - the centrality of the topic of his investigation - through emphasizing its importance (the importance of cement fitness), as opposed to what it leaves unaffected:

> *[makes changes: deletes "It is important to note that the total heat released during hydration is unaffected by the cement fineness." and adds "The cement fineness is very important to control the heat of hydration".] (Student 54, Camtasia transcript)*

Another example of content addition is extracted from the last draft of Student 63, where he claims the interest of the research community in the topic (Move 1, claiming centrality) and the need for a better understanding of the topic (Move 2, justification):

> ***Added*** → *The demand for statistical information about the Brazilian population has become an issue of considerable interest. The society and politicians need to understand the effects of policies on certain groups of the population. (Student 63, first to last draft modification)*

Content for Move 2 was added by Student 62 in order to highlight an existing problem that will be addressed in his study.

> ***Added*** → *However, the integration of that stream of population has not shown the results expected for the agricultural development's projects. Latino farmers are not well connected to USDA and other agricultural organizations. (Student 62, first to last draft modification)*

Student 27 added new content which carries the functions of move 3 – stating the purpose of current research, indicating the methodology used, summarizing the findings obtained, and outlining the structure of the paper.

> *[adds: "Within the process of analysis, this study used content analysis to find out the historical flow of Chicago downtown. And the paper is structured in accordance with the five stages of Florida's creative economy. The results show that, indeed, the Chicago downtown, although has a relatively short history, match Florida's theory."] (Student 27, Camtasia transcript)*

- Lexical-level modifications

A total of 38 (49%) modifications made at lexical level were found in 16 Camtasia recordings and a total of 121 (42%) – in all first to last drafts. It was apparent that some lexical changes were directly related to certain rhetorical moves, while others were more general. Therefore, lexical output modifications were divided into move-specific and non-move-specific. The move-specific lexical modifications (28 in Camtasia data and 62 in first

to last drafts) were changes that consisted of vocabulary clearly signaling a particular function of a move. For instance:

> *[changes the problematic red sentence to begin with "the importance of". deletes a part of the sentence. modifies more: "In fact, the importance of cement fineness has been elevated to the heat of hydration." (Student 54, Camtasia transcript)* – Move 2
>
> *The separation of enantiomers has been focused because of the impurities of chiral drug [1]. → Numerous efforts have been invested in the separation of enantiomers because of the impurities of chiral drug [1]. (Student 32 first to last draft modification)* – Move 1
>
> *[changes: "For example, the programmers have to explicitly managing threads in order to deal with things like deadlock avoidance, synchronization and mutual exclusion." --> "For example, one of problems is that the programmers have to explicitly managing threads in order to deal with things like deadlock avoidance, synchronization and mutual exclusion." reads more] (Student 64, Camtasia transcript)* – Move 2
>
> *But this is a batch solution (needs all measurements first) and also the resulting joint optimization is computationally complex. → However, there is one major limitation in that this batch solution needs all measurements first. Another problem is that the resulting joint optimization is computationally complex. (Student 30, first to last draft modification)* – Move 2
>
> *[Changes: "So it is possible to set downtown Chicago into Florida's five economic transitions" --> "So this paper assumes that it is possible ..." (Student 27, Camtasia transcript)* – Move 3
>
> *In this article, we study on how and to what extent UCD and usability evaluation are integrated. → The purpose of this study was to develop a new method of UCD and usability evaluation integration. (Student 28, first to last draft modification)* – Move 3

Non-move-specific lexical changes totaled 10 in Camtasia data and 59 in first to last drafts. They were more generic in nature, made to remediate inappropriate expression, informality, and lack of cohesion as well as to improve the clarity of ideas through paraphrasing, which can be seen in the following examples:

- Modified expression (1in Camtasia data and 18 in first to last drafts)

[changes: "Rosen (1993) also introduced a variance estimator for ordered frames in order to attend demands of Statistics Sweden." --> "Rosen (1993) also introduced a variance estimator for ordered frames in order to address the needs of users of official data."] (Student 30, Camtasia transcript)

In addition, Colombia is first in Latin America with more than 710 thousand tons of oil in 2006 (Fedepalma, 2007). → In addition, Colombia is first in Latin America producing more than 710 thousand tons of oil in 2006 (Fedepalma, 2007). (Student 44, first to last draft modification)

- More formal academic vocabulary (3 in Camtasia data and 11 in first to last drafts)

So… "but" is in uh, oral English, not very academic. uh... 'however', maybe? [replaces "but" with unfortunately", the deletes it and writes: ", However, it suffers from a problem that". Deletes again and writes: "However, one limitation is that"] (Student 30, Camtasia transcript)

Jinhan Cho (2001) talked about a spin self-assembly method as an alternative for making well-organized multilayer films in a very short process time. → Cho (2001) reported on a spin self-assembly method as an alternative for making well-organized multilayer films in a very short process time. (Student 29, first to last draft modification)

- Cohesion (4 in Camtasia data and 9 in first to last drafts)

[goes to the revision box. Highlights "However" and changes it to "Therefore"] (Student 65, Camtasia transcript)

The work will focus on the identification and optimizationof key parameters for producing desired core-shell polymer nanoparticles morphologies by RAFT microemulsion polymerizations with surf-CTAs. → Overall, this work focuses on the identification and optimizationof key parameters for producing desired core-shell polymer nanoparticles morphologies by RAFT microemulsion polymerizations with surf-CTAs. (Student 58, first to last draft modification)

- Paraphrasing (2 in Camtasia data and 21 in first to last drafts)

[The modification is: "In order to get better insight into these massive data, many visualization tools have been invented specifically to view biological data." --> "Many visualization tools have been invented specifically to view biological data inside these massive dataset."] (Student 43, Camtasia transcript)

Furthermore, the high concentration of metal particle in TDRP may also improve the reaction performance. → Furthermore, we expect that the high concentration of metal particle in TDRP may contribute to the improvement of the reaction performance. (Student 65, first to last draft modification)

- Grammar-level modifications

The participants made fewer changes at the level of grammar compared to other types of output modifications – 5 (7%) were identified in Camtasia data, and 14 (5%) in last student drafts. Grammatical modifications were mostly related to verb tenses and forms, and a few were related to subject-verb agreement and plural noun forms.

- Verb tenses and forms (4 in Camtasia data and 12 in first to last drafts)

  [replaces "can" with "will"] (Student 29, Camtasia transcript)

  [changes "is" to "was" in "The objective of this study is to monitor the formation of lipids in the mesocarp and its change in terms of composition of TG, vitamin E and carotenes in the oil palm fruits of E. guineensis from week 12 to 24 after anthesis in Tenera materials planted in the central area of Colombia".] (Student44, Camtasia transcript)

  This language design will hide the concurrency related details while still providing with much of the benefits of much of the benefits of multicore processing. → Our language design hides the concurrency related details while still providing with much of the benefits of multicore processing. (Student 64, first to last draft modification)

  As the world's top economic entity, America has went through a series of economic stages in a relatively short history. → As the world's top economic entity, America went through a series of economic changes in a relatively short time. (Student 27, first to last draft modification)

- Subject-verb agreement (1 in Camtasia data and 1in first to last drafts; the same modification)

  Currently, activated carbon adsorption are utilized and research mostly [10] [11]. → Currently, activated carbon adsorption is utilized and most commonly researched [10] [11]. (Student 65, first to last draft modification)

- Noun plural (1in first to last drafts)

  In this model, tungsten cones with different size are used to enhance the optical field. → In our model, tungsten cones with different sizes were used to enhance the optical field. (Student 40, first to last draft modification)

- Structure-level modifications

The Camtasia data consisted of 10 (13%) changes made at structural level, and 60 (21%) such changes were extracted when comparing the 16 participants' first and last drafts. These modifications were made either at sentence level or at paragraph level. Cases in which clauses were altered, split, or combined were classified as sentence-level modifications, and cases in which the position of sentences was changed to a different paragraph were considered paragraph-level modifications. For example:

- Sentence-level modifications (7 in Camtasia data and 50 in first to last drafts)

  [changes: "Traditional design of anti-infection surface includes mechanical mixture of antibacterial drugs and grafting of antibacterial drugs by chemical methods, but these methods are difficult to process and to control the concentration of antibacterial drugs on the surface (Rossella Fattori, Tommaso Piva, 2003)." --> "Traditional design of anti-infection surface includes mechanical mixture of antibacterial drugs and grafting of antibacterial drugs by chemical methods (Rossella Fattori, Tommaso Piva, 2003). However, these methods are difficult to process and to control the concentration of antibacterial drugs on the surface."] (Student 28, Camtasia transcript)

  This sentence should be changed. Maybe it should be added in the former sentence. [goes to the revision box. reads aloud changes: "However, the estimated value is much larger than the experimental results [4-7,4-11]. This probably because in their experiments the tips used were larger than the laser wavelength, hence they could not be treated simply as small dipoles." --> However, the estimated value is much larger than the experimental results [4-7,4-11], which is probably because in their experiments the tips used larger than the laser wavelength, hence they could not be treated simply as small dipoles."] I just put a single sentence into a complex one. (Student 40, Camtasia transcript)

  However, they are not always perfectly integrated. Many reasons, such as outsourcing of usability evaluation, limited time, and other things, can affect the integration. → However, they are not always perfectly integrated for many reasons, such as outsourcing of usability evaluation, limited time, etc. (Student 28, first to last draft modification)

  In the next section, we'll analyze the present situation in some industries where FDI plays an important role, and FDI in different areas is also under analysis in order to

be compared to the present research result about FDI's impact to economy. → In the next section, we discuss the environmental policies and laws executed from 1980s, in which FDI and economic development were mentioned. Then, we analyze the present situation in industries where FDI plays an important role, and we compare the FDI influence on different areas of the economy. (Student 53, first to last draft modification)

- Paragraph-level modifications (3 in Camtasia data and 10 in first to last drafts)

  [in short, modifications consist of changing the placement of ideas into a more logical order and reducing redundancy to concise sentences] (Student 30, Camtasia transcript)

  …structure, yeah, this is the structure. I can move it right before the last sentence. [cuts the sentence and places it later in the text.] (Student 61, Camtasia transcript)

  Moved up and modified→ *Jinhan Cho (2001) talked about a spin self-assembly method as an alternative for making well-organized multilayer films in a very short process time. → Cho (2001) reported on a spin self-assembly method as an alternative for making well-organized multilayer films in a very short process time. (Student 29, first to last draft modification)*

  Partly moved down and modified → *Moreover, it is also performed a comparison with other variance estimators that take into account a possible effect on bias due to systematic sampling with probability proportional to size. The numeric illustration is implemented using stochastic simulation so as to reproduce a pseudo-distribution of the estimators. → Then, we perform a comparison with other variance estimators that take into account a possible effect on bias due to systematic sampling with probability proportional to size. (Student 63, first to last draft modification)*

- Mechanics

  In addition, the participants modified their output at the level of mechanics.

Specifically, the format of in-text citations used by the participants to make reference to previous research was modified, and punctuation was changed in a few of instances.

- Citation format (6 in Camtasia data and 6 in first to last drafts)

  I'm just trying to put all the references in the same format [changes "(Sarndall, 1992)"
  --> "Sarndall (1992)" (Student 63, Camtasia transcript)

Peter A. Chiarelli (2001) found that the surface assembled by spin-coating method is more homogeneous than by dip-coating. → Chiarelli (2001) demonstrated the ability to control the coverage of polyelectrolytes spin-assembled into electrostatically bound multilayer films with monolayer thicknesses on the order of angstroms and described a technique that produces adsorption of alternating multilayers onto a glass substrate in controlled amounts. (Student 29, first to last draft modification)

Phillip Stalley (2007) made an experimental analysis around the topic of FDI and its effect as a double-edged sword to developing countries. → Stalley (2007) made an experimental analysis around the topic of FDI and its effect as a double-edged sword to developing countries. (Student 53, first to last draft modification)

- Punctuation (4 in first to last drafts)

In this work we use [1] to develop a KF-CS algorithm to causally reconstruct image sequences using MR data. → In this study, we use [1] to develop a KF-CS algorithm to causally reconstruct image sequences using MR data. (Student 30, first to last draft modification)

The genus Elaeis is compounded by two species; Elaeis guineensis and Elaeis oleifera. → The genus Elaeis is compounded by two species - Elaeis guineensis and Elaeis oleifera. (Student 44, first to last draft modification)

The palm fruit is composed of three parts, the exocarp, the mesocarp which contains the crude palm oil red used for consumption human and raw material for countless industrial processes, and endocarp that encloses the endosperm or almond which gets the palm oil used variously in the industry. → The palm fruit is composed of three parts: the exocarp, the mesocarp that contains the crude palm oil used for human consumption and raw material for countless industrial processes, and endocarp that encloses the endosperm or almond which makes the palm oil used variously in the industry. (Student 44, first to last draft modification)

The resulting assemblies are dynamic structures and crosslinking is required to maintain structural integrity. → The resulting assemblies are dynamic structures, and crosslinking is required to maintain structural integrity. (Student 58, first to last draft modification)

It should be mentioned, however, that students' output modifications were not always made at one level, but at multiple levels at the same time. In other words, some idea units in Camtasia transcripts and some sentences in first-last drafts contained modifications that combined changes in content, vocabulary, grammar, structure, or mechanics. Twelve such

units (18%) were found in Camtasia data, and 68 such sentences (34%) were found in first to last drafts. Randomly extracted examples are provided below:

> *[scrolls up in the revision box, fixes some articles, makes up for the cohesion, and adds: "Mesocarp is composed by tryglicerols (TG), carotenes and vitamins; TG, represent between 90 -95% of the total composition of lipids (Sundram et al., 2003), carotenes and, vitamin E during the developing process of the fruit."] (Student 44, Camtasia transcript)*
>
> *[goes to the revision box] I just delete the 'but', so... And then I add 'however', and it becomes a separate sentence. [changes: "Traditional design of anti-infection surface includes mechanical mixture of antibacterial drugs and grafting of antibacterial drugs by chemical methods, but these methods are difficult to process and to control the concentration of antibacterial drugs on the surface (Rossella Fattori, Tommaso Piva, 2003)." --> "Traditional designof anti-infection surface, which has attracted the attention of many researchers, includes mechanical mixture of antibacterial drugs and grafting of antibacterial drugs by chemical methods (Rossella Fattori, Tommaso Piva, 2003). However, these methods are difficult to process and to control the concentration of antibacterial drugs on the surface."] (Student 28, Camtasia transcript)*
>
> *Because MR data acquisition is sequential, the scan time (time to get enough data to accurately reconstruct one frame) is reduced if fewer measurements are needed for accurate reconstruction and hence there has been a lot of interest in the MRI community to use compressed sensing (CS) to do this [2],[3]. → Because MR data acquisition is sequential, the scan time (i.e. the time to get enough data to accurately reconstruct one frame) is reduced if fewer measurements are required for accurate reconstruction. Hence, there has been a lot of interest in the MRI community to use compressed sensing (CS) for this purpose [2],[3]. (Student 30, first to last draft modification)*
>
> *In 2007, the number of newly established foreign-invested enterprises is 37,888, a decrease of 8.69%. → In 2007, the number of newly established foreign-invested enterprises was 37,888, which is a decrease of 8.69% compared to last year. Nevertheless, the capital involved increased by 12.8%, amounting to 82.657 billions. (Student 53, first to last draft modification)*
>
> *Obviously, the enhancement is very importance for the processing. Nano-simulation processes offer a low cost, scalable alternative to laser beam design process. → Obviously, the enhancement is of great importance for the processing given that nano-simulation processes offer a low cost, scalable alternative to laser beam design process. (Student 40, first to last draft modification)*

5.4. Learning gains

Learners' interaction with IADE through its feedback appeared to enhance their learning gains. This claim is sustained by evidence obtained from 1 Likert-scale question, 5 survey questions, and from participants' scores on pre- and post-tests, which are summarized in Table 5.15.

Table 5.15. Overall evidence of learning gains

Data source	*N of participants*	*Question*	*Evidence for LLP*	*No evidence for LLP*
		Responses to:		
Survey data: Likert-scale q-ns	88	3: How well did you learn the moves?	98.86%	1.14%
Survey data: Yes/No & open-ended q-ns	83	2: Did you learn the moves? What makes you say that?	93.97%	6.03%
		3: What do you think helped you learn the moves the most?	92.77%	7.23%
		4: Would you say that you learned the moves as a result of having focused on them? Why?	93.98%	6.02%
		5: Would you say that you can transfer this knowledge to your writing? Why?	77.11%	22.89%
		7: Has the program helped you learn new words/expressions that signal particular moves/steps? How? Please provide examples of the words/expressions you learned for each move/step.	71.08%	28.92%
		Scores for:	t-test	
Pre-/post tests	104	104 pre-/post tests, task 1	p=.000	
		104 pre-/post tests, task 2	p=.000	
		104 pre-/post tests, task 1 and 2	p=.000	

5.4.1. Learner perceptions on learning gains

As shown in Figure 5.15, a majority of 79.54% (70 participants) who completed the Likert-scale questionnaire believed that they learned the moves 'well', 6.82% (6 participants) thought they learned them 'very well', and 12.5% (11 participants) – 'a little'. Only 1 participant (1.14%) chose 'not at all' when self-evaluating his learning gains.

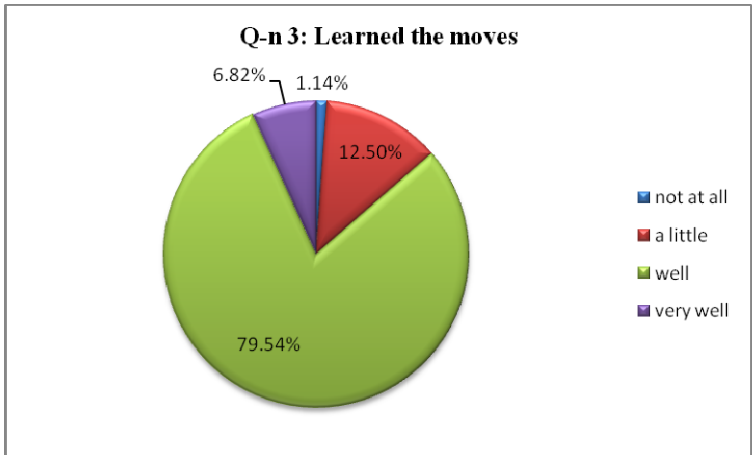


Figure 5.15. Results from Likert-scale question 3.

The responses to survey question 2 contained 78 'yes' answers (93.97%) with which the participants agreed that they learned the moves after having revised their work with IADE and 5 'no' answers (6.03%) formulated as "not completely" (1 student), "not really" (1 student), and "not very well" (3 students). The follow-up question about what made the students think that they learned or did not learn the moves was not addressed by all the participants – 21.69% (18 students) did not provide an answer, and 10.84% (9 students) provided irrelevant answers such as:

Yes. After the program analyzed, sometimes I have to (Student 38, survey question 2)

Yes, it's useful (Student 64, survey question 2)

Yes. I think the moves are very useful. (Student 32, survey question 2)

The 61.45% (51 students) who did address the follow-up to question 2 appeared to refer to themes similar to those that emerged from survey question 8 – causes of improvement and perceived or observed outcomes (see Section 5.1.2.1). The themes that rounded up in participants' perceptions here and are depicted in Figure 5.16 were causes of learning and perceived or observed outcomes. The participants thought that they learned the moves for one of two reasons - interaction with IADE through its feedback and Help Options (21.56% or 11 students) and their focus on discourse form (7.14% or 4 students). The perceived outcomes they mentioned was their enhanced understanding of the moves and their functions (58.87% or 29 students), and the observed outcome was improved quality of their final drafts (11.76% or 6 students). Also, like in the results obtained from survey question 8, the focus on form, enhanced understanding, and better quality of end products themes add to the evidence of the learning cycle presented in Section 5.2.2.

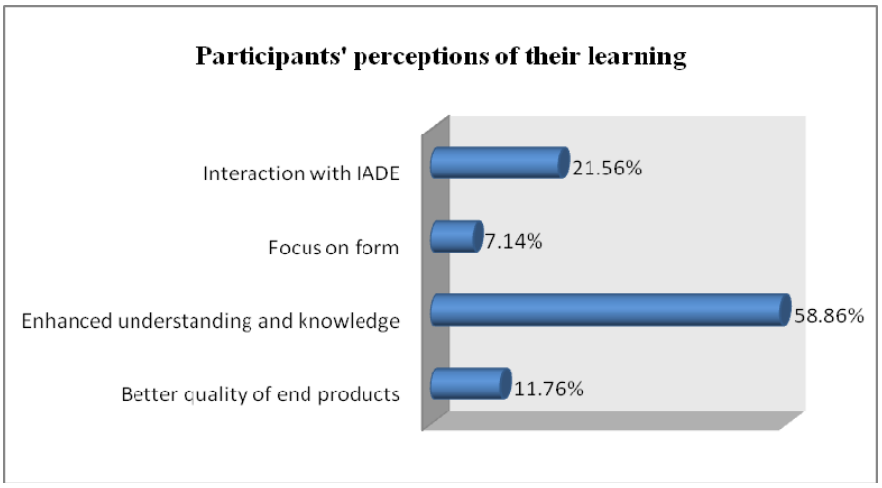


Figure 5.16. Results from survey question 2

- Interaction with IADE through its feedback and Help Options

  I had a plan that which sentence belongs to which move because the feedback give me different answer. (Student 44, survey question 2)

I learned in the class with the program (Student 48, survey question 2)

The tips and definitions remind me about these moves and made me learn. (Student 59, survey question 2)

I learned moves and its definitions from "Definitions" (Student 35, survey question 2)

I learn it from the definition part of IADE. *(Student 11, survey question 2)*

I learned because of the examples in the software of my discipline. (Student 10, survey question 2)

- Focus on form

because the program already highlighted the moves (Student 54, survey question 2)

I've learned a lot about the moves, focusing on their exact definition and the structure they should be organized. (Student 85, survey question 2)

I have tried to pay attention which move the part that I read is for when I review literatures. (Student 27, survey question 2)

While I was revising the introduction, I thought what is my sentences for. (Student 37, survey question 2)

- Enhanced understanding and knowledge of moves

I think at least I have a much better understanding. (Student 28, survey question 2)

Because I now know the definition of each move and its function of each move (Student 36, survey question 2)

Now I can distinguish which sentence belongs to which move and I know the structure of the introductions in my discipline (Student 50, survey question 2)

I know the function of each move in the introduction section. (Student 74, survey question 2)

I am able to write the introduction on my own with all the moves in it. I am able to recognize which sentence belongs to which move. (Student 33, survey question 2)

Because now I can determine these moves when reading research articles in my other courses. (Student 30, survey question 2)

- Better quality of end products

  My introduction is more organized. (Student 67, survey question 2)

  My moves made my paper very clear. (Student 61, survey question 2)

  My paper looks more professional because I learned these moves. (Student 41, survey question 2)

  I learnt the moves because I write what I intended to write and the decision of IADE *was good. (Student 31, survey question 2)*

  I could apply the moves in my research introduction. (Student 61, survey question 2)

  When I wrote my research paper, I introduced the problem very clearly. (Student 51, survey question 2)

The open-ended survey question 3 inquired more specifically about what the participants thought helped them learn the moves the most, and 92.77% of the responses (77 students) commented on a number of aspects of IADE's learning potential. Moreover, the responses once again formed themes that surfaced in other survey questions (see questions 2 and 8). As listed in Figure 5.17, easy access to the Help Options appears to have been most helpful. More precisely, definitions and accompanying examples of moves and steps were mentioned as most helpful by 26.51% (22 students) and the annotated corpus in the student's discipline by 18.07% (15 students). Also, 10.84% (9 students) believed that the color-coded feedback helped them learn the most because it drew their attention to segments of their text that needed improvement. The process of interacting with IADE in general was helpful to 16.87% (14 students). Some participants referred to a number of specific aspects of this interaction: the fact that interaction with IADE made them notice vocabulary indicative of a particular function of a move was mentioned by 6.02% (5 students), and that it made them focus on discourse form and see negative evidence by 4.82% (4 students) and 1.21% (1

student), respectively. Further, 8.43% (7 students) thought that the opportunity to practice and experiment with different ways to build rhetorical functions by resubmitting their modifications for automated analysis was a major factor that lead to the learning of moves. Interestingly, 7.23% (6 students) gave credit to classroom instruction as opposed to revising with IADE.

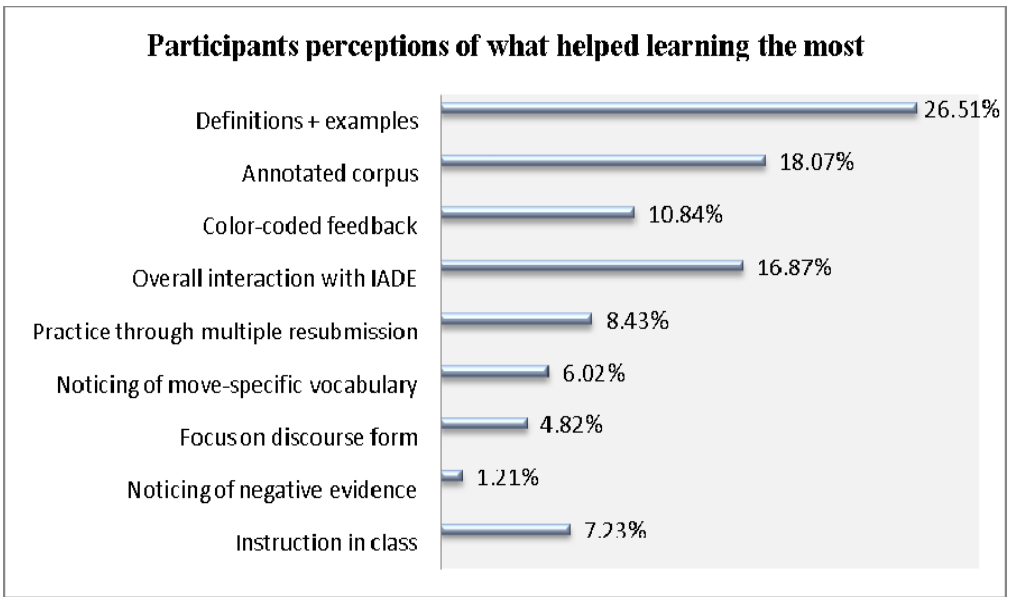


Figure 5.17. Results from survey question 3

Following are examples of participants' responses containing evidence of the themes that emerged in survey question 3:

- Definitions and examples of moves in Help Options

  The tutorial and the examples in each move. (Student 80, survey question 3)

  The notes given in help(Student 39, survey question 3)

  The definition part of IADE. *(Student 11, survey question 3)*

  The definitions and examples sentences. (Student 52, survey question 3)

  Class materials inside the program (Student 48, survey question 3)

The explanation helps me to learn the moves the most. (Student 40, survey question 3)

- Annotated corpus in Help Options

Read, read and read annotated corpus. With this I learned how to write steps in the moves. (Student 47, survey question 3)

The annotated corpus. (Student 35, survey question 3)

Working with the moves in several labeled introductions among my discipline was fundamental to learn the moves. (Student 8, survey question 3)

Reading related introductions in Help (Student 38, survey question 3)

The clear moves in the corpus and where it should be. (Student 85, survey question 3)

I learn the moves from the examples of the corpus (Student 50, survey question 3)

The corpus. It helped see clearly how to organize moves and how to express (Student 12, survey question 3)

- Color-coded feedback

Color-coding (Student 64, survey question 3)

Different colors to display differed moves (Student 20, survey question 3)

At first, I am not familiar with the moves, sometimes I have troubles in telling the difference between the moves. After using IADE*, the different colors can make the moves very easy to tell, and it helped me learn a lot. (Student 51, survey question 3)*

The sentences labeled by colors. (Student 28, survey question 3)

Highlighting the texts with colors according to the moves (Student 33, survey question 3)

To recognize them by color helped a lot. (Student 43, survey question 3)

- Practice trough multiple resubmission

I learn the moves from the revising process. (Student 44, survey question 3)

Analyze and revise (Student 62, survey question 3)

Usefully practice (Student 41, survey question 3)

To keep revising them and submitting and checking if it is good or not (Student 63, survey question 3)

Practicing and excising until the program shows it's good (Student 19, survey question 3)

Submit many times and know what is the reason for the answer. (Student 56, survey question 3)

- Overall interaction with IADE

Lecture's explains, reference materials, other examples connecting moves, and IADE. *(Student 34, survey question 3)*

Software provided by my instructor (Student 22, survey question 3)

I learned the steps of introduction with moves perfectly because of this program (Student 70, survey question 3)

The IADE *helped me learn the moves. (Student 55, survey question 3)*

The IADE *program. (Student 5, survey question 3)*

The software gave me exact idea on those moves (Student 74, survey question 3)

- Noticing of move-specific vocabulary

To find the key words about the moves (Student 53, survey question 3)

Key words and expressions and logical clue. (Student 27, survey question 3)

Keywords (Student 13, survey question 3)

Some key words can help me to find those moves. (Student 60, survey question 3)

When I learned the key words I learned the moves and steps (Student 37, survey question 3)

That I figured out which words and how to properly choose the words in each move. (Student 1, survey question 3)

- Focus on discourse form

Because I considered so much where are my moves, I learned how to write the introduction logically (Student 57, survey question 3)

I looked too much at the logic and structure. (Student 73, survey question 3)

Every time I focused on how I organized my ideas, so I think that's how I learnt the moves pretty well. (Student 46, survey question 3)

I saw how my organization changed when I changed my moves and I saw what's good and when (Student 7, survey question 3)

- Noticing of negative evidence

It makes me see what's wrong and clean every move. (Student 4, survey question 3)

- Class instruction

My teacher (Student 42, survey question 3)

The gradual information that was taught in class and application of this information while doing the assignments helped me a lot in learning the moves. (Student 31, survey question 3)

The course (Student 76, survey question 3)

My teacher's course-work and her presentation. (Student 86, survey question 3)

The clear presentation and discussion in the class. (Student 30, survey question 3)

Class activities provided by instructor (Student 29, survey question 3)

Considering the theoretically supported importance of focus on form and its hypothesized role in this study, survey question 4 asked whether the participants thought they learned the moves as a result of having focused on the discourse form and why they thought so. The majority of respondents, i.e. 93.98% (78 students) said 'yes'; 2.4% (2 student) were not sure, and 3.61% (3 students) said 'no'. It must be pointed out here that the open-ended "Why?" question turned out to be poorly formulated judging by the answers it elicited. Specifically, 62.65% (52 students) did not provide an answer to the follow-up question at all. The responses of those who did answer were not all relevant to the question - 22.89% (19

students) answered relatively appropriately, and 14.46% (12 students) answered inappropriately, for instance:

> *Yes. It makes me read more. (Student 38, survey question 4)*
>
> *Yes, before I had written the introduction copying the structure of other papers (Student 28, survey question 4)*
>
> *They are not actually the same things I think (Student 12, survey question 4)*
>
> *Yes. Maybe not all. it seems that there is something difference in our field (Student 87, survey question 4)*
>
> *Yes, I learn the moves because I was focus in them. I should be able to do an introduction after the explanation. (Student 47, survey question 4)*

Although the evidence gained from this open-ended question cannot be claimed as strong, the few ideas that appeared in the appropriate answers present justifiable interest because they are supportive of some stages of the learning cycle described in Section 5.2.2. (also see Figures 5.5 and 5.10). In other words, it seems that the students realized that they noticed negative evidence, modified their output, and enhanced their understanding and knowledge of the moves because they had focused on the form of their introduction discourse. This can be seen in the following examples:

- Focus on discourse form leads to noticing of negative evidence and output modification:

  > *It made me see my weakness part of my writing. (Student 31, survey question 4)*
  >
  > *Yes. I focused on the moves to revise my draft. If I lack one move, I add some sentences represent this move. If I have too much of one move, I combine some sentences or delete some sentences to cut this move shorter. From this procedure, I learned what kind of sentences cannot express the moves correctly and how to revise them into explicit moves (Student 50, survey question 4)*

Yes. If I haven't focused on them, I can't even recognize them and see where my structure is not so good. (Student 27, survey question 4)

Yes, if I focus on my moves I know how I can put a certain bad sentence in a way that I want others to know which moves I'm in. (Student 53, survey question 4)

Yes, more focus on them make me change and write the article clearly. (Student 41, survey question 4)

- Focus on discourse form leads to enhanced understanding:

Yes, because focus on my moves helped me not only to recognize the moves but to understand them. (Student 43, survey question 4)

Yes, identify my own sentence make me understand the moves more clearly. (Student 37, survey question 4)

Yes. What caught my attention I think is the first step of learning. (Student 54, survey question 4)

Yes. Focusing on them helps me thinking and learning. (Student 63, survey question 4)

I may still have difficulties to identify them, but I think I understand and I know what I know because I payed a lot of attention to moves. (Student 75, survey question 4)

Yes, because only by focusing we are able to identify and make sense of it. (Student 39, survey question 4)

Yes, I would. The focusing on them thought me to realize the differences of the moves. (Student 22, survey question 4)

Yes, first, I have to focus on them then I can learn them. (Student 11, survey question 4)
Yes, as I have a better understanding of it after having focused on them. (Student 5, survey question 4)

Seeking evidence of participants' confidence in the newly acquired knowledge of Introduction moves, survey question 5 asked the participants whether they thought they could transfer what they learned to their actual writing. Most of the participants felt that they could do that (77.11% or 64 students); 16.87% (14 students) were not so confident, saying "I don't

know", "I am not sure", "I will try", "possibly", "maybe", "I hope so", or "kind of", and 6.02% (5 students) did not think they were ready to freely apply their declarative knowledge in order to produce well-rounded introductions, explaining that they may need more practice:

> *Too little exercise. (Student 58, survey question 4)*
>
> *Not yet, I still try to figure it out and write some more, then transfer my knowledge. (Student 4, survey question 4)*
>
> *It's too time consuming and too complex, so I need to use the software more and maybe write introduction not for only one paper. (Student 15, survey question 4)*
>
> *My knowledge is too fresh. It take time and writing before I transfer it. (Student 38, survey question 4)*
>
> *Because when I writing a paper, I still focus on the content too much. (Student 36, survey question 4)*

Those who were more optimistic about their learning gains as applied to their writing skills mostly justified that by naming the actions that they would take to ensure the transfer of knowledge, among which were: focusing on and analyzing their moves (64.06% or 41 students) which is also indirect evidence for focus on form, matching whether the functional meaning of their moves is successfully rendered through vocabulary (21.88% or 14 students), and comparing them with published professional texts (14.06% or 9 students). Consider the following examples:

- Focus on and analysis of moves

> *I know that there exists a better way to write a scientific article, and I will pay attention to that. I will pay a lot of attention to every move I write. (Student 82, survey question 4)*
>
> *I didn't know there're moves before, but now I will focus on them as much as I can, I mean I'll try to identify them in my paper and do what I did for this paper to make it better (Student 68, survey question 4)*

When I write articles right now, I would think of the knowledge and look sentence by sentence if I applied everything I learned. (Student 63, survey question 4)

I will focus a lot on organizing the moves of my introductions. (Student 85, survey question 4)

I can transfer because I have analyzed how I write and organize the moves. (Student 49, survey question 4)

Definitely yes! I have a clear idea about in what I should focus and how to judge moves when I do an introduction. (Student 47, survey question 4)

After I learn this knowledge, last week, when I have to write a paper, I use this and I found that it is so easy for me to introduce the problems. (Student 51, survey question 4)

- Matching vocabulary with functional meaning

I will pay attention to the words and the meaning I want to present in the sentences. (Student 57, survey question 4)

I have attainted how to make sure what I say is what I mean. (Student 40, survey question 4)

I know keywords in sentence are very important to make my paper understood. I used them and I will use them. (Student 61, survey question 4)

I can easily frame the introduction by using the right words for right moves. (Student 39, survey question 4)

Because when I writing, I focus more on the meaning I want to expressed, but for the meaning I will choose key-words. (Student 35, survey question 4)

Actually I did it. Because recognized in what part of the introduction my concept should be included and how it need to be expressed. (Student 8, survey question 4)

- Comparison with other professional texts

Because I will compare my work with others. (Student 54, survey question 4)

I will see that in my day to day work and writing and also in journals. (Student 70, survey question 4)

Right now, I know I will look for sure at my corpus for good models to make sure my is good too. (Student 22, survey question 4)

I will use this software to set up good habit of writing. (Student 51, survey question 4)

Because it is very useful if I look at how other papers are written, I think I'll compare with corpus again and again. (Student 34, survey question 4)

Since the results of the previously conducted pilot study indicated that the participants noticed move-specific phraseology while revising with IADE, the survey in this study included a question (# 7) about noticing and possibly learning new words or expressions that signal particular moves/steps, to which 71.08% (59 students) answered "yes" and 28.92% (24 students) answered "no", "not so much", "nope", "not quite", "not really", "I don't think so", or "No, it doesn't help me a lot in this function" (Student 55, survey question 7). Of the participants who chose to say "yes", 40.68% (24 students) explained that what made them notice certain vocabulary was accessing sample moves in IADE's annotated corpus and in the explanations of moves provided in the Help Options (27.12% or 16 students) as well as multiple modifications of their own output (13.54% or 8 students). For instance:

- Accessing examples in IADE's Help Options

About the new words and expressions, I learned it from the corpus, which is provided by this program. This corpus of my area did help me to grasp the words and expressions, which are very useful in my own paper writing. (Student 86, survey question 4)

Yes, in annotated corpus (Student 62, survey question 4)

Using the program with several different articles showed that there are certain words for each move. (Student 27, survey question 4)

I learned new words for moves from the definition part. (Student 11, survey question 4)
From the examples in definition and corpus, I learn some new words/expressions in Move Three. (Student 44, survey question 4)

- Output modifications

When I changed the words to express my meaning, sometime the move changed. (Student 67, survey question 4)

Yes, I tried the following sentence it works for 3 moves: "I am a good person. - Move1; However, some people think I am not a good person. - Move2; In this paper, I proposed I am a good person." - Move 3 (Student 82, survey question 4)

If I change certain words, the move of the sentence can possibly be changed. Especially for move 3, when I wrote a certain sentence of methodology, it cannot recognize it. Only when I use the word "we", or "in this paper", etc, it can indicate that this paragraph belong to move 3. But I think it is awkward to use these words in each sentence. (Student 72, survey question 4)

Other 22.03% (13 students) both commented on how they noticed move-specific phraseology and provided specific examples of lexical items that they thought signaled certain moves, and 37.29% (22 students) only listed examples of words or expressions they had learned while revising with IADE. Some of those are quoted below:

"In the next section, the more detailed methodology is described." is label as move 3 but "More detailed methodology is described next." is not (Student 57, survey question 4)

For example, Move 1: interest, importance, emphasize: Move 2: no/ little work has been reported on; If this tendency were accounted for / true, results could differ substantially; Move 3: In my research; Our experiment focus on; This paper present that (Student 30, survey question 4)

Like using "issue" instead of "topic" for example. The using of "issue" made the sentence in move 2 instead of move 1 when using "topic" (Student 31, survey question 4)

When I worked with the corpus, I learned many expressions. For example, I can replace "describe" with "examine, assert, dispute, or claim". (Student 35, survey question 4)

Move3-the goal, the purpose; Move 2- however, limits of this research (Student 25, survey question 4)

For example how to use modal verbs (could/would/should) to give some justification meaning to the sentence, or suggesting things and not assuring them. (Student 8, survey question 4)

For example, I learned new ways to express centrality without showing explicit the importance of the topic, expression as: recent years. (Student 71, survey question 4)

5.4.2. Learning gains measured through tests

As described in Chapter 4, Section 4.6.1.4, the pre- and post-test scores were analyzed to measure participants' learning gains after they revised their research article introduction drafts with IADE. Since the tests consisted of two tasks requiring the identification of moves and steps (the first task consisting of de-contextualized sentences and the second of an entire introduction text), descriptive statistics and t-tests were calculated for scores on each task and for combined scores on both tasks. Tables 5.16a and 5.16b show the results for task one, for task two, and for all the test scores. All the post-test means are higher than the pre-test means, and the probability values of .000 reflect significant leaning gains on the part of the participants.

Table 5.16a. Descriptive statistics for learning gains (based on pre- and post test scores for both tasks)

		Mean	*N*	*Std. Deviation*	*Std. Error Mean*
Pair 1	Post_test1	29.65	104	1.909	.187
	Pre_test1	20.50	104	4.942	.485
Pair 2	Post_test2	30.16	104	1.596	.156
	Pre_test2	21.05	104	6.750	.662
Pair 3	Post_test_all	59.82	104	2.988	.293
	Pre_test_all	41.55	104	10.161	.996

Table 5.16b. T-test results (based on pre- and post test scores for both tasks)

		Paired Differences					*t*	*df*	*Sig. (2-tailed)*
		Mean	*Std. Dev.*	*Std. Error Mean*	*95% Confidence Interval of the Difference*				
					Lower	*Upper*			
Pair 1	Post_test1 - Pre_test1	9.15	5.11	.50	8.16	10.15	18.26	103	.000
Pair 2	Post_test2 - Pre_test2	9.11	6.76	.66	7.80	10.43	13.75	103	.000
Pair 3	Post_test_all- Pre_test_all	18.27	10.35	1.01	16.25	20.28	17.99	103	.000

5.5. Effect of practice on improvement

Evidence that learners' practice in generating new output and modifying their original discourse through multiple resubmissions to IADE contributed to learning and improvement in the rhetorical quality of their discourse was found in survey question 9. The relation between the frequency of draft submissions and final draft scores was also examined. Table 5.17 summarizes the overall findings from these data sources.

Table 5.17. Overall evidence of the role of practice through multiple resubmissions

Data source	*N of participants*	*Data*	*Evidence for LLP*	*No evidence for LLP*
Survey data: Likert-scale q-ns	88	Responses to: Q-n 15: Did you get enough writing practice with the program?	100%	0%
Survey data: Yes/No & open-ended q-ns	83	Q-n 9: Did revising and resubmitting your drafts help you improve? Why?	92.77%	7.23%
IADE's database; IADE's automated evaluation	105 105	Submission frequency & scores for: 1,173 submissions 210 first and last drafts	$r = .286$	

5.5.1. Learner perceptions on revision and resubmission

All the participants felt that they had sufficient writing practice opportunity. In scale question 15, 21.59% (19 students) indicated that they had "a lot" of practice, 51.14% (45 students) thought it was sufficient, and 21.59% (24 students) chose "a little" (see Figure 5.18).

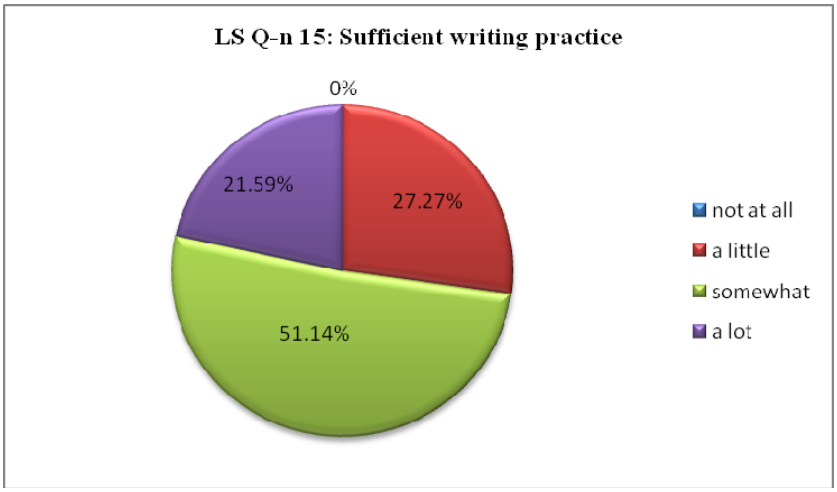


Figure 5.18. Results from Likert-scale question 15

According to participants' responses to survey question 9, practice in the form of multiple revisions and resubmissions stimulated by IADE's feedback was considered helpful for improvement by 92.77% (77 students). Four participants (4.82%) found it of no help, and 2 participants (2.4%) were not sure whether it was helpful or not. This is how the two students explained their uncertainty:

> *I don't know if I improved because of revising and resubmitting but I do know that I tried to pay more attention when I wrote. (Student 23, survey question 9)*
>
> *Not sure, no one except the program gave me the feedback of my draft. (Student 34, survey question 9)*

The “no” responses to this question were not accompanied by any explanations; however, the positive answers were more informative. As shown in Figure 5.19, a number of participants (28.57% or 22 students) explained that the opportunity for practice, which they had due to revision based on iterative feedback and multiple resubmissions, actually helped them consolidate the knowledge about the moves that they were acquiring. Here, a potential connection may be made between participants’ perceptions of consolidated knowledge and their confidence that they would be able to transfer the newly acquired knowledge to their further academic writing practice (see Section 5.4.1, survey question 5). Also, although not very strong, this perception of knowledge consolidation may serve as indirect evidence speaking in support of the role of practice for transforming declarative knowledge (which the participants were exposed to in class instruction) into automated skill, as sustained by the Skill Acquisition Theory.

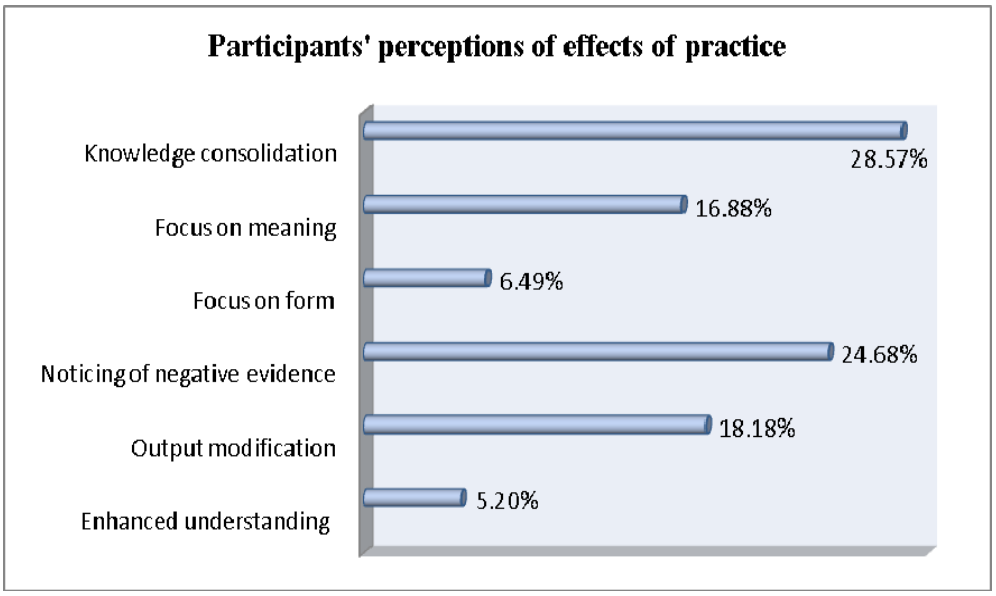


Figure 5.19. Results from survey question 9

Among the reasons why other participants perceived revising and resubmitting as helpful were the elements of the learning cycle discussed in Section 5.2.2, although those elements do not appear in any sequence in this part of the data. Some (6.49% or 5 students) believed that resubmitting helped them focus on their discourse form; some (24.68% or 19 students) thought the more they revised and resubmitted, the more negative evidence they could identify; some (18.18% or 14 students) claimed to have been able to make more output modifications that were eventually successful; and others (5.2% or 4 students) explained that this type of practice lead to an enhanced understanding of the rhetorical moves. Following are random examples of participants' explanatory follow-ups to survey question 9:

- Consolidation of knowledge

  When you revised it, you have to read it again and this makes you know for sure what you think you know a little. (Student 38, survey question 9)

  Basically I did revise my draft many times and I'm sure I learned more and more about moves every time I resubmit. To learn the moves well is not enough to submit one time. (Student 45, survey question 9)

  Practice is always helpful. For me writing is like something else I learn to do. You do it many times and you finally do very well. Like experiments. (Student 43, survey question 9)

  Revising and resubmitting your drafts helped me improve. I already knew about moves from my teacher, but I didn't know well until I revise my own moves (Student 70, survey question 9)

  Revising always helps. This time it also helped me because I submit again and l when it's correct I have good knowledge finally and I'll write correct the next time. (Student 5, survey question 9)

  Revising is always helping me to learn new things. Like the bycicle, I fall and stand up many times and I finally learn to ride the bysicle. So it's the same with my introduction, I revise and revise until I have no problem. (Student 86, survey question 9)

- Focus on meaning

  Revising made me go back and think what I have wrote is really what I meant or not. (Student 41, survey question 9)

  I know I did think a lot about if my moves sound with the right meaning and how I want every time I revised. (Student 6, survey question 9)

  I start to know how to formulate my moves clear. (Student 4, survey question 9)

  because I didn't express what I want rightly in the beginning, but then I revise and resubmit and finally make what I mean clear for the software. (Student 7, survey question 9)

  I don't think the original one is too good because my move sound different than I mean. (Student 55, survey question 9)
  At least the moves seem not bad because the program recognize the meaning of my moves at the end. (Student 76, survey question 9)

- Focus on form

  help me to know the structure of my drafts. (Student 1, survey question 9)

  it makes me pay attention to the structure. (Student 85, survey question 9)

  of course this process helps the learning procedure because you see how your moves get better organiz (Student 28, survey question 9)

  I do not have much time in revising it again and again, but when I did few times, I looked at how my introduction is outlined in moves. (Student 67, survey question 9)

  Because when I submit I can look to see if my moves are logical and follow the order in my field. I revise and submit and check again. (Student 82, survey question 9)

- Noticing of negative evidence

  Through the revising, I learned what I need to improve (Student 52, survey question 9)

  I realized my research draft was deficient of the sentences and moves of general previous researches. That is, my research will not give readers my research purpose and results and after all will not be assessed by them. (Student 34, survey question 9)

  Revising my drafts helped me to find a lot of problems in my paper and know the shortage of my writing. (Student 59, survey question 9)

Because I can note what were mistakes. (Student 47, survey question 9)

Because I came to know what errors I made. (Student 33, survey question 9)

Because I usually made many mistakes in the first drafts. For example, sometimes I use move 2 style to express move 3 ideas. (Student 25, survey question 9)

- Output modification

I can modify the language and structure. (Student 27, survey question 9)

Because with revising, I can make some improvement by myself such as elongation of whole introduction. (Student 56, survey question 9)

Resubmitting was really helpful as through it I changed and improved each move. (Student 29, survey question 9)

I wrote very short introduction in my first draft. After I got IADE *evaluation, I added stuff and improved my introduction (Student 22, survey question 9)*

The statistics in IADE *after submission really helped me to revise with good changes and improve my introduction. (Student 70, survey question 9)*

I add some move 2. (Student 24, survey question 9)

- Enhanced understanding

Revising and resubmitting help me understand the expression of each move (Student 57, survey question 9)

Because revision makes the writer to look from others point of view to his writing (Student 31, survey question 9)

I have written imaging different moves, but after I ran this software several times I came to know what are the specific moves used in my department and how much of the percentage I need to include in the introduction. (Student 35, survey question 9)

During the revising, I can understand why some words are often used in the research articles. (Student 11, survey question 9)

A new theme mentioned by 16.88% (13 students), also given in Figure 5.23, is focus on meaning, which in fact presents great interest for the research question about the Meaning

Focus quality discussed in Chapter 6. In this regard, the participants made it clear that, because they had the chance to resubmit their work for automated analysis many times, they found themselves thinking more and more about the meaning they were trying to convey.

5.5.2. Relation between practice and improvement

Practice in the context of this analysis was looked at in terms of frequency of draft submission. The drafts of 105 participants were submitted 1,173 times over a period of 6 days, as recorded in IADE's database. Table 5.18 summarizes the frequencies with which the participants submitted their drafts and the number of their submissions. The submission numbers are divided into frequency bands for a more compact presentation of the results. As it can be seen in Table 5.17, the lowest submission frequency band was 3-5 times (17.14% or 18 students). A great many of participants (44.76% or 47 students) revised and resubmitted their drafts 6-10 times and another 27.62% (29 students) – 11 to 20 times. A few students submitted their work more than 30 times (10.47% or 11 students).

Table 5.18. Draft submission frequency bands and numbers of participants

Frequency band	*# of participants*
3-5	17.14% (18)
6-10	44.76% (47)
11-20	27.62% (29)
21-30	7.62% (8)
35	0.95% (1)
43	0.95% (1)
65	0.95% (1)
Total	100% (105)

In addition, as Table 5.19 shows, most of submissions were done within 2-3 days with an average of 43 submissions per participant; 12.38% (13 students) finished their work within one day, submitting their drafts for automated analysis on average 10 times; 20.95%

(22 students) continued working for 3 days with 10 submissions on average; and a few other participants took up to 6 days to revise, although those who took 6 days did not make many submissions.

Table 5.19. Number of submissions by time period

Submission period (days)	*# of participants*	*# of submissions*	*Avg # of submissions per participant*
1	12.38% (13)	11.68% (137)	10
2	57.14% (60)	61.55% (722)	43
3	20.95% (22)	18.84% (221)	10
4	5.71% (6)	5.63% (66)	11
5	1.91% (2)	1.87% (22)	11
6	1.91% (2)	0.43% (5)	2
Total	100% (105)	100% (1,173)	14.5

To examine the relationship between practice and overall draft improvement, the score differences for each move and length were summed up and used to calculate the Pearson product-moment correlation. A statistically significant correlation was found at the 95% level of confidence with a correlation coefficient $r = .286$ (N=105), which means that practice through multiple resubmission of revised drafts was positively related to overall improvement of students' writing.

In order to find whether the frequency of submission, or the practice variable, had an impact on the improvement of individual aspects of students' drafts, One-Way ANOVA tests on three score difference groups (score difference = 2, score difference = 1, and score difference = 0) were conducted for Move 1, Move 2, Move 3, and draft length, respectively. Variances for these groups satisfied the constant variance assumption, except for length, and ANOVA tests, with a significance level set at .05, were performed to compare the group

means for each aspect at a time. Tables 5.20a and 5.20b present the respective descriptive statistics.

Table 5.20a. Relationship between practice and improvement on moves

Draft aspect	*Score difference group*	*N*	*Mean*	*Std. Dev.*
Move 1				
	0	45	9.93	8.971
	1	45	11.76	10.377
	2	14	13.85	9.929
	Total	104	11.25	9.727
Move 2				
	0	38	9.03	5.838
	1	40	12.23	12.050
	2	27	12.63	10.024
	Total	105	11.17	9.714
Move 3				
	0	20	11.10	8.522
	1	55	9.38	6.940
	2	30	14.50	13.549
	Total	105	11.17	9.714

Table 5.20b. Relationship between practice and improvement on length

Draft aspect	*Score difference group*	*N*	*Mean*	*Std. Dev.*
Length				
	0	73	10.59	8.244
	1	25	11.96	8.502
	2	7	14.43	22.619
	Total	105	11.17	9.714

The ANOVA results showed that there were no significant differences in the number of submissions across the three score difference groups for Move 1 ($F = .975$, $p = .381$), Move 2 ($F = 1.480$, $p = .233$), and draft length ($F = .602$, $p = .549$). Move 3 was the only aspect for which more encouraging evidence was found. For this move, the number of submissions was highest for the group with the highest score difference and therefore with the highest improvement. Although the results showed no significant difference among the

groups ($F = 2.788$, $p = .066$) at 95% confidence level, it was statistically significant at 90% confidence level. This result sheds light on a finding introduced earlier in Section 5.3.2, according to which most 'noticeable' and 'considerable' improvement was made for Move 3. It is reasonable to assume, then, that better progress on Move 3 could be related to practice through more frequent resubmissions.

The fact that no statistical significance was found for the relationship between individual moves and draft length and the submission frequency may be due to the small sample (as shown in Tables 5.19 a and 5.19b, N ranged between 7 and 73), which resulted in low power of ANOVA. However, the fact that a strong relation was found between the submission frequency and overall improvement may be explained in view of the nature of resubmitted revisions. As the qualitative analysis of improvement revealed, the students acted upon the negative evidence they noticed, which was not only related to discourse form, but also to discourse meaning (see Section 5.3.4). Consequently, they made revisions which improved the rhetorical quality of the entire text, not only its form.

5.6. Effect of modified interaction on improvement

Whether modified interaction contributed to learning and improvement or not was examined with the help of 1 Likert-scale question and 3 survey questions. In addition, the data consisting of participants' access to IADE's Help Options, which was stored in the program's database, was analyzed for statistical significance when correlated with their draft scores. The database stored automatically tracked participants' clicks on the four available options: definitions and examples of moves and steps, discipline-related statistics for steps, the annotated corpus, and the revision tips.

Table 5.21. Overall evidence of the role of modified interaction through IADE's Help Options

Data source	*Nr of participants*	*Data*	*Evidence*[4] *for LLP*	*No evidence for LLP*
		Responses to:		
Survey data: Likert-scale q-ns	88	Q-n 4: Did you use IADE's Help Options?	92.05%	7.95%
Survey data: Yes/No & open-ended q-ns	83	Q-n 10: Do you think it is helpful that IADE offers Help Options? Why?	83.13%	16.87%
		Q-n 11: During the process of writing and revising your drafts, did you access any Help Options? Which Help Options did you use most often? Why?	91.57%	8.43%
		Q-n 12: Did you access the annotated corpus in the Help Options? Why?	65.06%	34.94%
		Frequency of clicks & scores for:		
IADE's database; IADE's automated evaluation of drafts	105	1,312 clicks on Help Options	93.34%	6.66% *r* = .-177
	105	210 first and last drafts		

5.6.1. Learner perceptions on helpfulness of modified interaction

According to participants' responses to survey question 11, 8.43% (7 students) did not access any help and 91.57% (76 students) did. Very similar results were obtained from Likert-scale question 4 but with a few more details (see Figure 5.20). About a quarter of participants (23.86% or 21 students) indicated that they made use of IADE's Help Options "a

[4] For frequency of clicks, % indicates the percent of participants who clicked on IADE's Help Options.

lot", 26.14% (23 students) - "somewhat", 44.32% (37 students) - "a little", and 7.95% (7 students) did not use them at all.

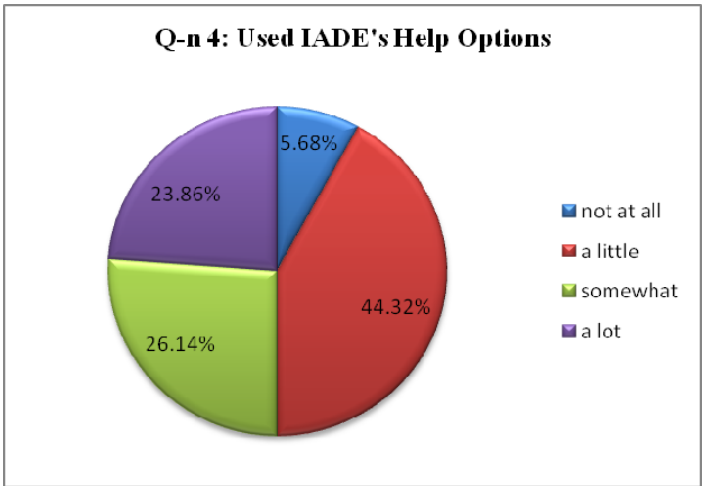


Figure 5.20. Results of Likert-scale question 4

The same type of information was captured in IADE's database. Of 105 participants, 6.66% (7 students) indeed did not access the Help Options at all, while the remaining 93.34% (98 students) interacted with IADE for the purpose of draft revision. The overall numbers of clicks on all four options per participant are given in Table 5.22. Apparently, most of the students' modified interaction activity fell into one of three bands for frequency of clicks: 1-5 (24.76%), 6-10 (28.57%), or 11-20 (23.81%). A few participants elicited help more than 20 times, one of them doing it 87 times.

Further, Table 5.23 shows that a total of 1,312 clicks were recorded, the overall average per participant being about 12 clicks. It appears that the explanation of moves and steps were most frequently consulted by learners (742 times), followed by the examples in the annotated corpus (243 times). Disciplinary statistics and revision tips were accessed less frequently (178 and 149 times, respectively). Similar were participants' answers to the open-

ended part of survey question 11, in which they named what Help Options they used most frequently. Specifically, of 76 participants who indicated that they worked with IADE's Help options, 53.94% (41 students) mentioned the definitions and examples of moves and steps, 27.63% (21 students) mentioned the annotated corpus, 7.9% (6 students) mentioned the disciplinary statistics describing the distribution of steps, and 10.53% (8 students) mentioned the tips offered for more effective revision.

Table 5.22. Frequency bands for clicks on all Help Options and numbers of participants

Total per participant	
Frequency band	*# of participants*
0	6.66% (7)
1-5	24.76% (26)
6-10	28.57% (30)
11-20	23.81% (25)
21-30	8.57% (9)
31-40	5.71% (6)
51-60	0.96% (1)
81-90	0.96% (1)

Table 5.23. Number of clicks on IADE's Help Options per participant

	Total # of clicks	*Average # of clicks*
Definitions	742	7.06
Annotated corpus	243	2.31
Disciplinary statistics	178	1.7
Revision tips	149	1.42
Total	1312	12.49

With more details, Table 5.24 and Figure 5.21 present the frequency bands for participants' access to each of these help choices. Between 12.38% and 28.57% of them did not access some of the Help Options. The most prominent frequency band for all Help Options is 1-5; consequently, most of the participants engaged in modified interaction with each option at least 1 to 5 times.

Table 5.24. Frequency bands for clicks on each Help Options and numbers of participants

Definitions		*Annotated corpus*		*Disciplinary statistics*		*Revision tips*	
Freq. band	# of participants	Freq. band	# of participants	Freq. band	# of participants	Freq. band	# of participants
0	12.38% (13)	0	20.95% (22)	0	24.76% (26)	0	28.57% (30)
1-5	52.38% (55)	1-5	67.62% (71)	1-5	72.38% (76)	1-5	68.57% (72)
6-10	17.14% (18)	6-10	10.48% (11)	6-10	2.86% (3)	6-10	1.91% (2)
11-20	9.52% (10)	11-20	0.96% (1)			11-20	0.96% (1)
21-30	4.76% (5)						
31-40	1.91% (2)						
51-60	0.96% (1)						
61-70	0.96% (1)						
Total	100% (105)		100% (105)		100% (105)		100% (105)

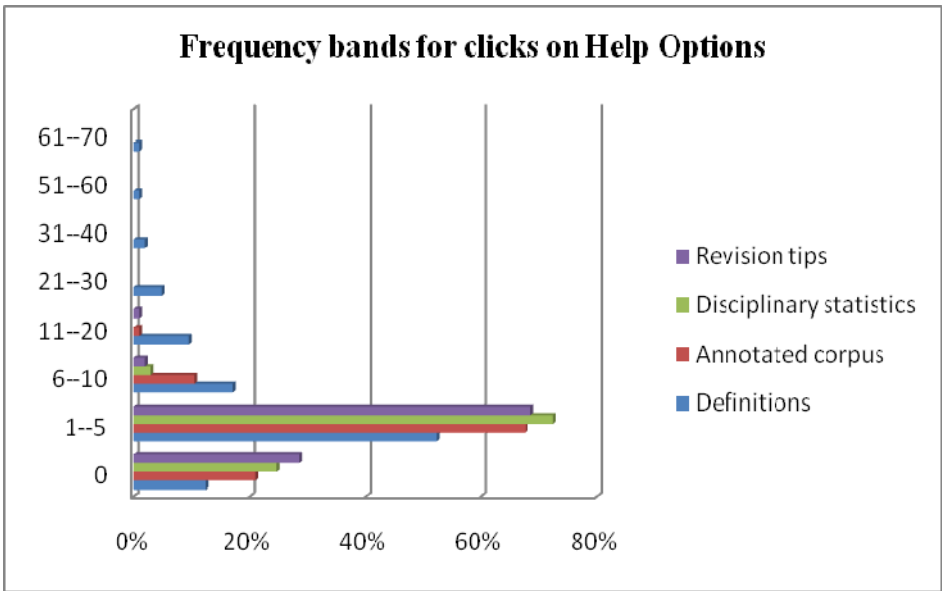


Figure 5.21. Frequency bands for clicks on each of the Help Options

Survey question 10 attempted to gain an insight into learners' perceptions of whether it is helpful to have Help Options in a program like IADE and why. Of 83 participants, 83.3% (69 students) thought the availability of such a feature was helpful; 3.61% (3 students) thought it wasn't; 4.82% (4 students) were not sure, saying "maybe'; and 8.43% (7 students) did not have an opinion, explaining that they did not access IADE's Help Options. The students who were not sure whether such options would help or not did not provide any

explanation for their uncertainty, while 72.46% (50 students) of those who were confident of their positive answers provided reasoning that formed three themes: knowledge consolidation, revision strategy development, and practicality. The knowledge consolidation category contained 48% of the positive answers (24 students), which explained that having Help Options readily available would be helpful because they would be a source that would remind them of what they've learned in class or would help them verify their new knowledge and hypotheses. The revision strategy development category consisted of 16% of answers (8 students), which made reference to the revision tips that guide learners through the revision process, helping them develop more successful revision strategies. The third theme, practicality, was formed by 36% of answers (18 students) about the convenience that the Help Options would ensure. These three themes are shown in Figure 5.22 and exemplified with participants' responses below:

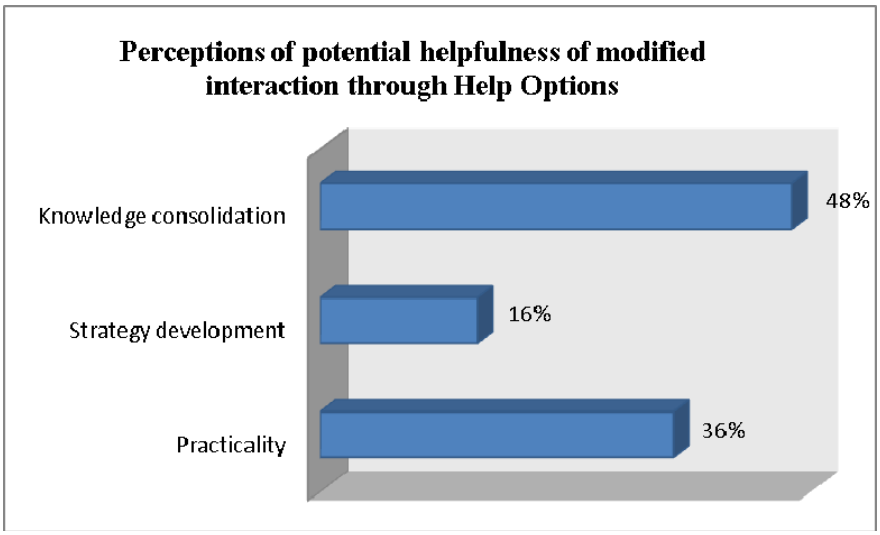


Figure 5.22. Results from survey question 10

- Knowledge consolidation

  very helpful because we cannot remember so many moves and steps. (Student 21, survey question 10)

I can review some knowledge. (Student 54, survey question 10)

I can get help when I'm not so sure about the individual move. (Student 85, survey question 10)

sometimes I cannot remember the moves so I can turn to the help option (Student 65, survey question 10)

It would help me to understand moves by describing and defining moves (Student 35, survey question 10)

Every option would help me better learn about some points easily ignored. (Student 30, survey question 10)

- Strategy development

  or else even I know what the problem is, I don't know how to revise it. *(Student 10, survey question 10)*

  It gives me guides when I am lost. (Student 63, survey question 10)

  It can supply some good tips for how to revise. (Student 51, survey question 10)

  This gave some hints to go on and improve the writing (Student 31, survey question 10)

  because you have a reference of how to do it well. (Student 8, survey question 10)

  You can have access to some articles in your field and check the percent of the move. This can help you when you write. (Student 47, survey question 10)

- Practicality

  Without the Help Options, the program would become much less convenience. (Student 59, survey question 10)

  It makes it simple to understand the IADE *better. (Student 55 survey question 10)*

  It could help you to better use the program. (Student 42, survey question 10) Otherwise I may get confused of the recommendation of the software. (Student 72, survey question 10)

  Help options are important for any program in different reasons for users. (Student 52, survey question 10)

I can easily search for help there. (Student 11, survey question 10)

It is also important to note that the knowledge consolidation and revision strategy development themes appeared as well in the second open-ended survey question 11, which asked the participants why they accessed IADE's Help Options. As shown in Figure 5.23, of 76 participants who indicated that they had used IADE's help, 53.95% (41 students) said that they needed that help because they did not have solid knowledge of the moves and steps, and 14.47% (11 students) worked with IADE's help to get some guidance in their revision process. The other two themes that surfaced in participants' answers to this question were follow-up on feedback (26.32% or 20 students) and vocabulary search (5.26% or 4 students). In the former, participants said that they looked up Help Options when they needed assistance in order to address the negative feedback, and in the latter a few participants indicated that they used the help to find appropriate move-specific phraseology. See examples below:

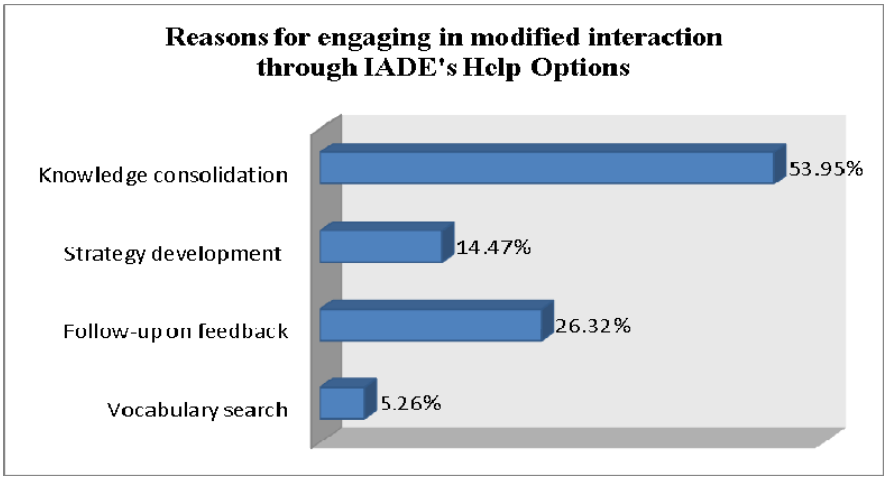


Figure 5.23. Results from open-ended survey question 11

- Knowledge consolidation

  I used the definition and examples in the "Help Options". Sometimes, I forgot the meaning of the moves and steps. So, the "Help Options" can help me to understand them. (Student 28, survey question 11)

  because I cannot remember every move. (Student 36, survey question 11)

  I can't remember each section in each move exactly. (Student 3, survey question 11)

  I may see some tips on what each move mean to get some idea on those moves which I am not sure. (Student 74, survey question 11)

  The meaning of move. Sometime I forgot the definitions of them. (Student 67, survey question 11)

  I forgot the characteristics of each move. (Student 54, survey question 11)

  Because sometimes I am confused about different steps and their function and help options can tell me about what I need. (Student 13, survey question 11)

- Strategy development

  Sometimes I don't know how to revise my drafts, and I will access 'Help Options'. *(Student 57, survey question 11)*

  I tried to do what the tips say, and after a couple times I know what to do and how to revise good. (Student 83, survey question 11)

  Need to see tips to know what to do first and then next. (Student 64, survey question 11)

  I would access the Revision Tips to revise the draft. (Student 51, survey question 11)

  I could just click on Revision Tips and get a solution. (Student 12, survey question 11)

  When I start, I didn't know how to revise except mistakes in my grammar, but the Revision Tips help me make a plan of steps. (Student 7, survey question 11)

- Follow-up on negative feedback

  When I revised my drafts, I confused some sentences that I suppose it is move 3; however, IADE *appeared it is move 2. Thus, I use "Help Options." (Student 40, survey question 11)*

to find out what I have missed in the writing. (Student 15 survey question 11)

I need to know reason for the negative valuation by the IADE. *(Student 66, survey question 11)*

when I want to improve move 1 which is only 40% of the whole paper. (Student 93, survey question 11)

when I can't figure out what IADE *suggested, I will look for help. (Student 17, survey question 11)*

because I have some questions about the wrong color. (Student 9, survey question 11)

- Move-specific vocabulary search

I used examples because I am not very sure about the steps, and I want to find some expression to write my own draft (Student 44, survey question 11)

I was trying to find synonymous for steps in order to avoid repetition (Student 16 survey question 11)

the annotated corpus to make sure of the language I used. (Student 31, survey question 11)

I take account into the color of words to make different moves more clear. And I'll also use the "View an annotated corpus" to view similar examples with different words. (Student 102, survey question 11)

As said in Q6, I check the keywords. (Student 37, survey question 11)

because I need the examples of expressions of the moves. (Student 19, survey question 11)

Given that class instruction followed a corpus-based approach and that it included work with annotated corpora, survey question 12 referred to participants' use of the annotated corpus. Answering this question, 65.06% (54 students) said that they accessed the annotated corpus in IADE's Help Options and 34.94% (29 students) didn't. Of those who did not access the annotated corpus, only the following few explained why, mentioning reasons such as not having noticed the availability of the annotated corpus in IADE, having worked

with it before in class, or the fact that the corpus was either not in their discipline or that the texts were on topics too different from their research area:

> *I didn't notice it. (Student 63, survey question 11)*
>
> *I did not notice that. (Student 104, survey question 11)*
>
> *I do not have time to check them. (Student 64, survey question 11)*
>
> *because the corpus of my field is not put into the database yet. (Student 2, survey question 11)*
>
> *I have checked those papers in class Callisto. (Student 63, survey question 11)*
>
> *I do not when I can use it. Because we have already analysis this articles on class before. I don't think it's very useful and helpful for my study and research because it's about other areas of my field. (Student 19, survey question 11)*
>
> *because the information not really matches my filed of research. (Student 70, survey question 11)*
>
> *the articles in the corpus in not very related to what I plan to write about. (Student 72, survey question 11)*

Unlike the negative answers, more positive answers were accompanied by explanations than not – 33.33% did not follow up on the open-ended "why?" question, and 66.67% (36 students) provided a number of reasons, which are listed in Figure 5.24. Two of the reasons provided by these participants coincide with the themes identified in survey questions 11 and 12 – knowledge consolidation (19.44% or 7 students) and vocabulary search (27.78% or 10 students). Another prominent theme is that of accessing the annotated corpus to find examples of moves and steps, which contains 30.56% of explanations (11 students). In addition, 22.22% (8 students) looked up the corpus to compare their drafts with published texts in their discipline.

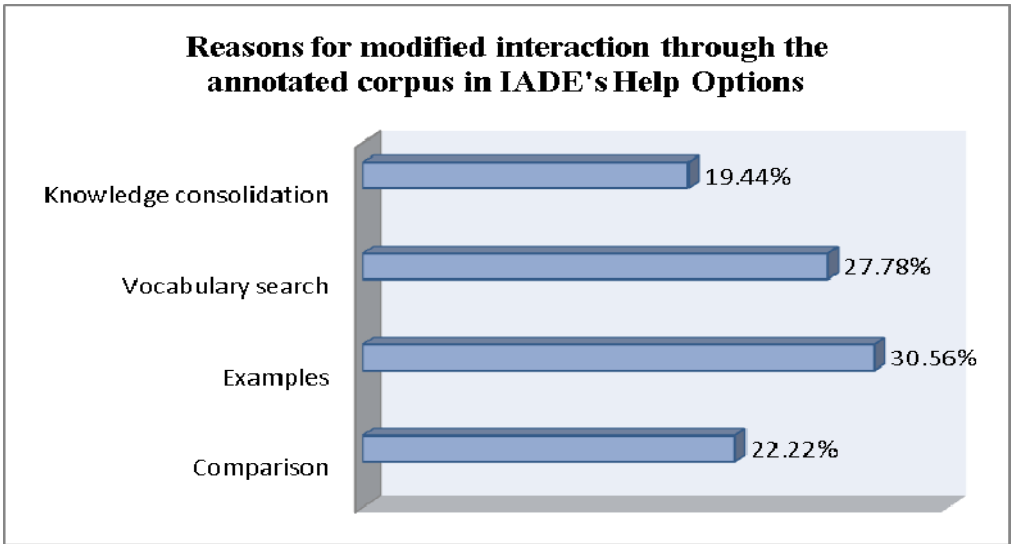


Figure 5.24. Results from open-ended survey question 12

- Knowledge consolidation

  When I want to see more about the moves (Student 21, survey question 12)

  It is helpful of knowing move step, words, and construction of sentences. (Student 34, survey question 12)

  Full content helps me to understand well (Student 82, survey question 12)

  Sometime I am not very sure about the move. So I will access the annotated corpus to learn something. (Student 71, survey question 12)

  To recall the moves. (Student 6, survey question 12)

- Vocabulary search

  it helps me with the key words. I regard the corpus as reference. (Student 53, survey question 12)

  because I do not know how to write a step, and I want to see some examples of expressions. (Student 50, survey question 12)

  To check for some suitable expressions (Student 31, survey question 12)

  To check words or sentences for Move 2. (Student 30, survey question 12)
  It is helpful and provides some useful expression to me. (Student 20, survey question 12)

- Examples

  Sometimes; when I confused some moves, I tried to find some examples to figure out what I want. (Student 41, survey question 12)

  To see examples. (Student 26, survey question 12)

  I needed to see more examples of moves for understanding. (Student 48, survey question 12)

  I need those corpus to see examples how moves and steps are in my field. (Student 91, survey question 12)

  I used many times during the work session to look at examples of steps when I'm not sure. (Student 49, survey question 12)

  It's really good for examples. You can click on sentence and see the step. (Student 24, survey question 12)

- Comparison with disciplinary texts

  to compare with articles in my field (Student 43, survey question 12)

  I wanted to have an idea about how my field evaluate the introduction. (Student 97, survey question 12)

  I accessed to check other articles moves and see if my introduction is like them. (Student 79, survey question 12)

  they help me to know classical papers' method to organize their introduction and their expression methods. (Student 86, survey question 12)

  Because it tell me what other article look like. (Student 10, survey question 12)

  I did because I wanted to know how the other introductions of my field are. (Student 18, survey question 12)

The prevalence of the knowledge consolidation theme explains the high frequency of participants' access to the definitions in IADE's Help Options, where they could see explanations of moves and steps that reminded them of rhetorical functions and possibly even helped them learn. Based on this evidence, which is also supported by the data from

think-aloud/Camtasia, observations, and interview transcripts, it can be inferred that modified interaction through IADE's help with definitions was a factor that played an important role in the enhanced understanding element of the learning cycle discovered during the participants' revision process (see Section 5.2.2). Similarly, the participants looking for examples and move-specific vocabulary as well as their need to compare their work with that of professionals in their field may be connected to their clicking on the annotated texts, which had the second highest frequency. Although rarely mentioned in responses to open-ended questions, the disciplinary statistics given in IADE's Help Options may have also been consulted for the purpose of comparison since this type of help was the next in frequency after the annotated corpus. Finally, participants' access to the revision tips among IADE's Help Options may have contributed to their development of new revision strategies and, in particular, actions taken to follow-up on the negative feedback. These connections are graphically presented in the summative description of participants' modified interaction through each of IADE's Help Options given in Figure 5.25.

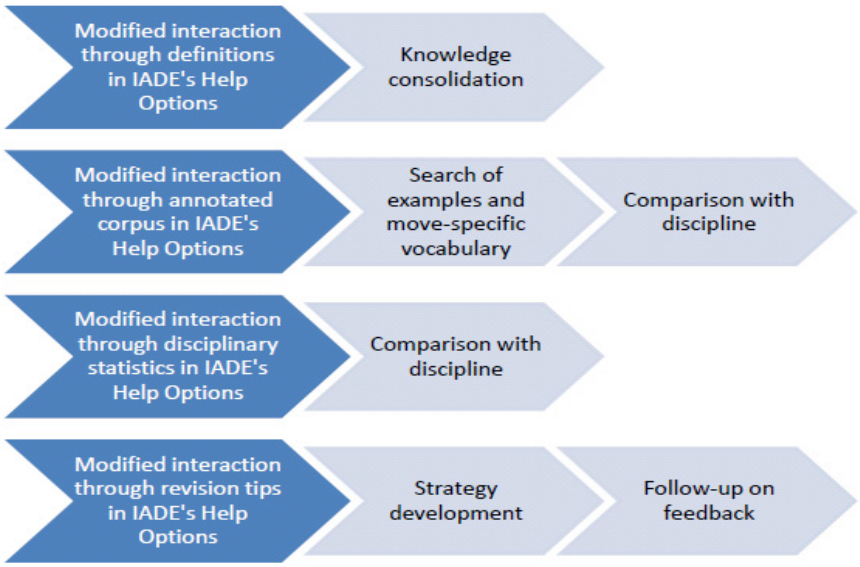


Figure 5.25. Modified interaction through IADE's Help Options

5.6.2. Relation between modified interaction and improvement

To find whether there was a relationship between modified interaction and overall draft improvement, the Pearson product-moment correlation was calculated. No statistically significant correlation between the total number of clicks on Help Options and the overall draft improvement scores was found (r = .-177, N=105). However, a possible reason for this result may be the small sample size. Also, after analyzing the data by separating it into individual moves and length modules, it became clear that the patterns were conflicting and, therefore, it is not surprising that no significant correlation was found. Figure 5.26 depicts these conflicting patterns.

One-Way ANOVA tests, with a significance level set at .05, were performed to examine the effect of modified interaction on the improvement of Move 1, Move 2, Move 3, and draft length. Table 5.25 contains the respective descriptive statistics. The ANOVA results revealed no statistical differences. It was interesting, however, to see a small exception in the case of Move 1. For this move, weak evidence was found (F = 2.712, p = .071), which allows for the inference that the group with the highest improvement accessed the Help Options the least number of times. This may mean that the students who made the highest progress on Move 1 had acquired better declarative knowledge of this move and, thus, did not have much need to consult the supporting materials provided IADE's Help Options. Or, perhaps this was a move that was least difficult. These assumptions, however, are not supported by the evidence in this study.

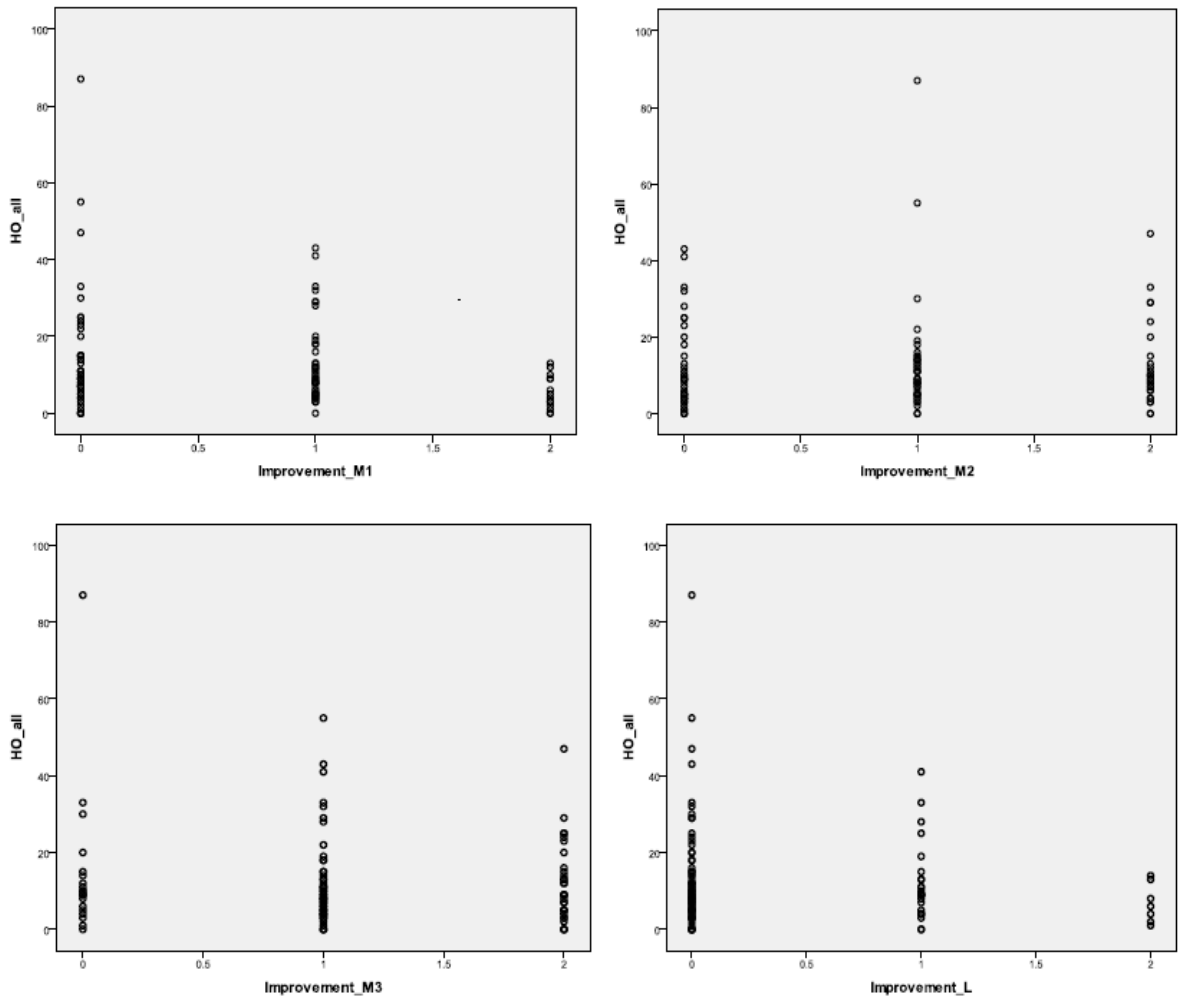


Figure 5.26. Patterns between frequency of access on Help Options and improvement on Move 1, Move 2, Move 3, and draft length.

Table 5.25. Relationship between modified interaction and improvement on moves and length

Draft aspect	*Score difference group*	*N*	*Mean*	*Std. Dev.*
Move 1				
	0	45	14.56	16.026
	1	45	12.58	10.409
	2	14	5.50	4.398
	Total	104	12.48	12.921
Move 2				
	0	38	12.05	11.524
	1	40	12.93	15.234
	2	27	12.48	11.126
	Total	105	12.50	12.860

Table 5.25. Relationship between modified interaction and improvement on moves and length (continued)

Draft aspect	*Score difference group*	*N*	*Mean*	*Std. Dev.*
Move 3				
	0	20	14.80	18.998
	1	55	11.85	11.385
	2	30	12.13	10.520
	Total	105	12.50	12.860
Length				
	0	73	13.22	14.095
	1	25	11.96	10.143
	2	7	6.86	5.113
	Total	105	11.17	12.860

5.7. Overall evaluation of Language Learning Potential

This chapter presented the findings obtained from both quantitative and qualitative data gathered in this study, containing strong evidence that IADE and its automated feedback hold strong Language Learning Potential. Evidence of Language Learning Potential was sought through analyses that focused on students' noticing of and focus on discourse form, remediation indicative of improvement, learning gains, revision practice, and modified interaction.

Multiple sources were analyzed to obtain the necessary empirical evidence of focus on form. The quantitative data considered for this purpose were 6 Likert-scale questions and 12 survey questions. Based on a response rate of 83.81% (88 out of 105 participants), the answers to the Likert-scale questions indicated that the Language Learning Potential quality of IADE is relatively strong. To allow for consistent interpretations, the scale answers were converted to evaluative comments; i. e., "not at all" was interpreted as poor evidence of the given CALL quality, "a little" – as weak, "much" – as good, and "a lot/very much" – as excellent. This data revealed positive findings in that 52% of respondents perceived the LLP

quality of IADE as being excellent and 30% as good. Negative evidence amounted to only 18% (see Figure 5.27).

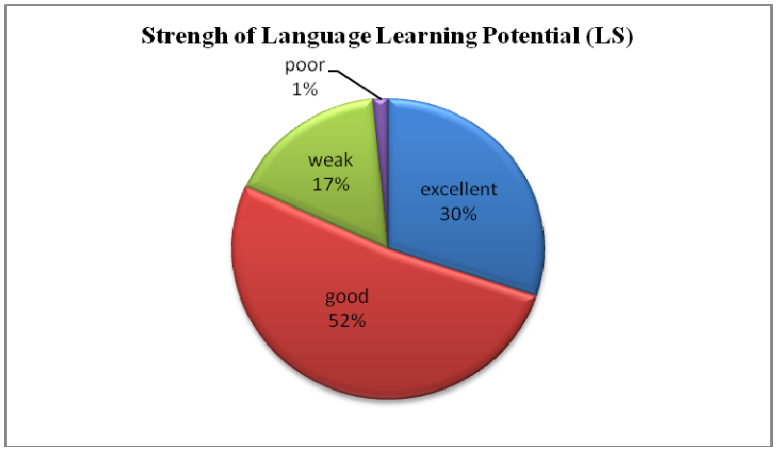


Figure 5.27. Strength of Language Learning Potential in Likert-scale data (N=88)

As detailed in Table 5.26 and in Figure 5.28, excellent evidence of different aspects of LLP ranges from 40.91% to 6.82%, good – from 79.52% to 26.14%, weak – from 44.32% to 6.8%, and poor – from 5.68% to 0%. Table 5.26 contains direct evidence of focus on discourse form, which is mostly good and excellent (47.7% and 44.2%, respectively), as well as additional evidence of LLP, which is also to a large extent positive. Specifically, a considerable number of the respondents to the scale questions selected "much" or "very much" when referring to the degree to which they noticed negative evidence in their own writing, noticed vocabulary indicative of a particular move, engaged in modified interaction, learned the moves, and improved their RA Introduction writing skills.

Table 5.26. Evidence of aspects of Language Learning Potential in Likert-scale data (N=88)

	Not at all (poor)	*A little (weak)*	*Much (good)*	*Very much (excellent)*
Focused on discourse form	1.1%	6.8%	47.7%	44.3%
Noticed negative evidence	1.14%	7.95%	50.00%	40.91%
Noticed vocabulary indicative of moves	0.00%	9.09%	52.27%	38.64%

Table 5.26. Evidence of aspects of Language Learning Potential (Likert-scale) (continued)

	Not at all (poor)	*A little (weak)*	*Much (good)*	*Very much (excellent)*
Engaged in modified interaction	5.68%	44.32%	26.14%	23.86%
Learned the moves	1.14%	12.50%	79.54%	6.82%
Improved RA Intro writing skills	0.00%	19.32%	54.54%	26.14%
Average	1.52%	16.66%	51.70%	30.12%

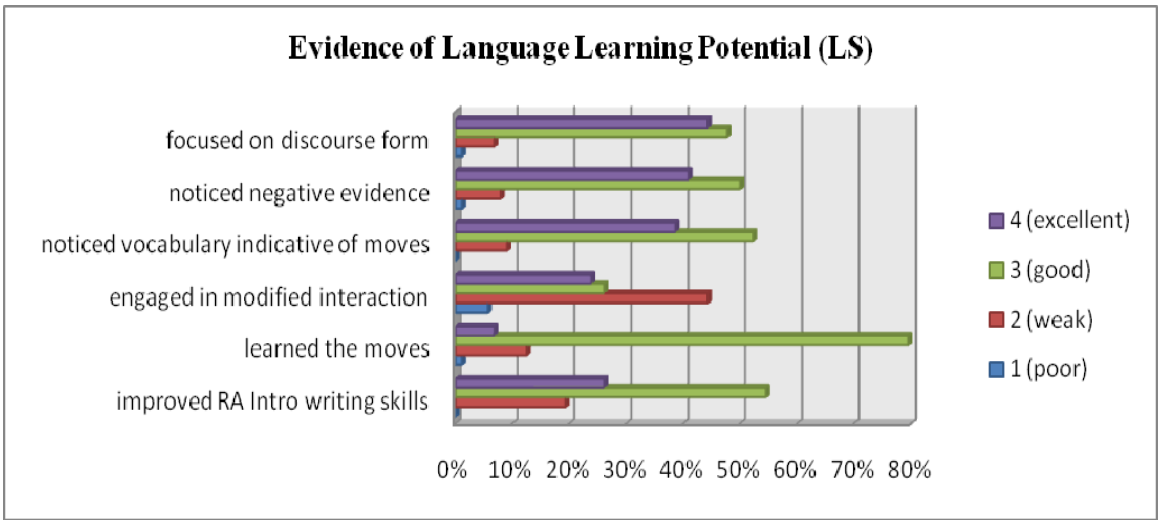


Figure 5.28. Evidence of aspects of Language Learning Potential in Likert-scale data (N=88)

The survey also revealed that the participants' perceptions were indicative of strong Language Learning Potential. Figure 5.29 shows that, on average, positive evidence amounted to 87%, and 13% of participants provided no evidence of this quality.

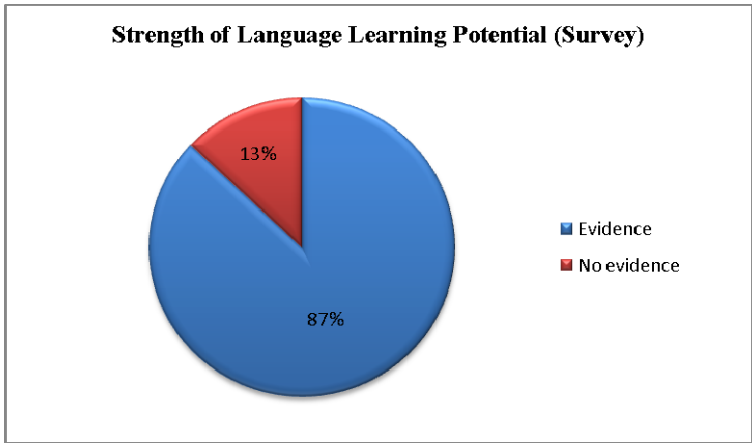


Figure 5.29. Strength of Language Learning Potential in Yes/No survey data (N=83)

Moreover, Table 5.27 and Figure 5.30 list a summary of specific findings obtained from participants' survey responses. The overall percentage of evidence speaking in favor of the LLP quality is 87%, the highest number in the range being 100% for focus on discourse form.

Table 5.27. Evidence of LLP in open-ended survey data (N=83)

	Evidence	*No evidence*
Focused on discourse form	100.00%	0.00%
Noticed peculiarities of discourse form	89.16%	10.84%
Improved RA Introduction writing skill	92.80%	7.20%
Learned the moves	93.97%	6.03%
Noted aspects that helped learning	92.77%	7.23%
Learned the moves as a result of focus on discourse form	93.98%	6.02%
Could transfer new knowledge to writing	77.11%	22.89%
Learned move-specific vocabulary	71.08%	28.92%
Found revision practice helpful	92.77%	7.23%
Found Help Options helpful	83.13%	16.87%
Engaged in modified interaction through HO	91.57%	8.43%
Engaged in modified interaction with annotated corpus	65.06%	34.97%
Average	87%	13%

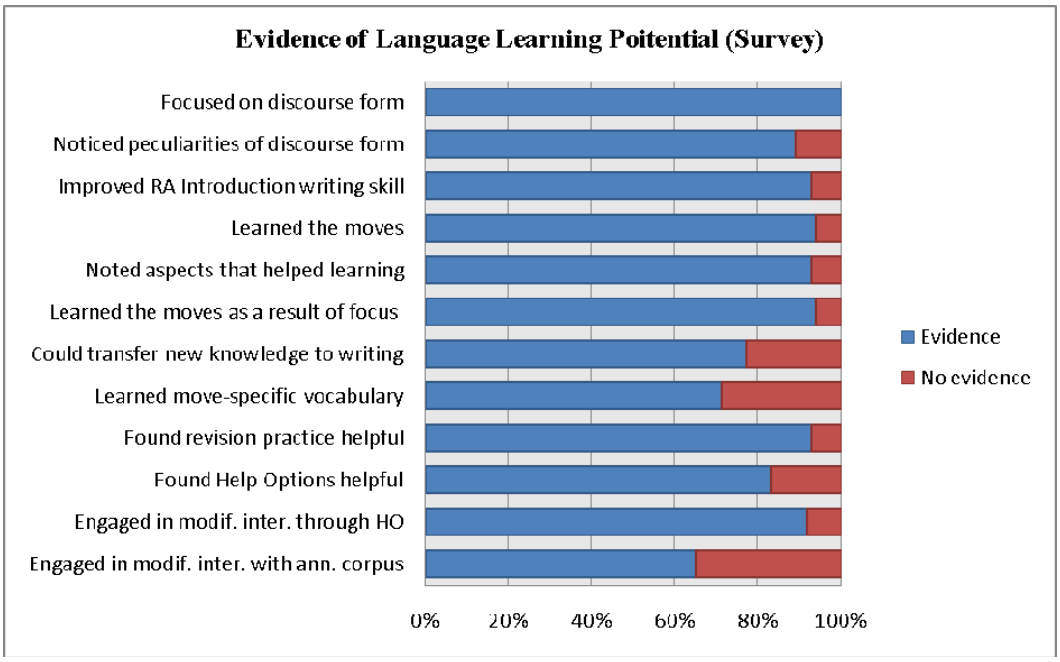


Figure 5.30. Evidence of aspects of Language Learning Potential in open-ended survey data (N=83)

Although engagement in modified interaction with the annotated corpus had the lowest 65.06%, it should not be considered as a weak piece of evidence since modified interaction in general through IADE's Help Options ranked much higher at 91.57%. Percentages for participants' perceptions regarding other aspects related to learning and the helpfulness of the program's features were largely positive as well.

In support of these findings, which reflect the students' perspective on the learning potential of IADE and its feedback, introspective and observational data (think-aloud protocols, interviews, observations, screen-capturing) indicated that learning occurred through intra-personal interaction that took the form of a cycle. This learning cycle consisted of four stages: focus on form, noticing of negative evidence, enhanced understanding of rhetorical moves, and output modification. Learners appeared to go through this cycle sequentially and iteratively, as a result revising and improving their written output.

Evidence of improvement was obtained by converging multiple data, both quantitative and qualitative. The analysis of scores based on IADE's evaluation not only showed significant improvement on individual moves and length from first to last draft (p =.000), but also revealed that 39.5% of improvement was noticeable and 18.5% was considerable. Limitations in IADE's analyzing techniques may be a possible explanation of why 35.2% were classified as needing no improvement. Similar analysis of scores for 32 first and last drafts rated by human raters showed higher percentages for degrees of improvement – 54.38% for noticeable improvement and 31.25% for considerable improvement. Moreover, manual analysis of output modifications made in these drafts identified 285 changes (18 changes per participants on average) that revealed improved quality of their academic

discourse. Modifications were made mainly at the levels of vocabulary, content, and structure and less in grammar and mechanics.

Furthermore, participants' pre/post test scores showed significant learning gains (p=.000 on t-tests), which may be a result of the learning cycle they appeared to engage in, particularly of the modified interaction through IADE's Help Options that was followed by output modification. According to the data captured in IADE's database, 93.34% of all participants elicited the program's help, having clicked on Help Options 1,132 times. Participants' open-ended survey responses contained explanations of why the Help Options were helpful. Although no statistically significant relationship between modified interaction and overall draft improvement was found, the Help Options seemed to have helped the learners to: consolidate new knowledge by accessing move/step definitions; search for move-specific phraseology in the annotated texts and compare them with their own drafts; and develop revision strategies, following up on the received automated feedback. In addition, the opportunity to practice through multiple resubmissions of drafts, which were submitted with a total of 1,173 times, was perceived by the participants as beneficial for knowledge consolidation and was found to be related to improvement.

To sum up, the major findings presented in this chapter, which support the conclusion that IADE and its automated feedback posses a strong Language Learning Potential quality, are:

- Participants noticed and focused on the RA Introduction discourse forms;
- Participants improved the rhetorical development of their RA Introduction discourse from first to last draft;
- Participants' interaction with IADE enhanced their learning gains;

- Participants' practice in generating new and/or modifying their original discourse through multiple resubmissions to IADE upon the reception of its automated feedback contributed to improvement in the rhetorical quality of their discourse;
- Participants' engagement in modified interaction through IADE's Help Options contributed to learning.

5.8. Chapter summary

This chapter presented results obtained through multiple analyses of different quantitative and qualitative data, which were converged in order to obtain evidence of IADE's Language Learning Potential. Responses to yes/no, open-ended, and scale questions were triangulated, as were the transcripts of think-aloud protocols, observations, and semi-structured interviews. In addition, juxtaposing automated and human ratings allowed for assessing the outcomes, or participants' work products, and comparing pre- and post-test scores helped evaluate their learning gains. Quantitative measures and perceptive and introspective insights complement each other in that the former provide a ground for generalizations and the latter provide an understanding of the process during which learning to write occurred. The overall findings warrant the conclusive claim that IADE and its automated feedback hold strong Language Learning Potential. This potential is realized through learners' noticing of and focus on discourse form, which generates a learning cycle that includes noticing of negative evidence, enhanced understanding, and output modification. Learners' engagement in this cycle can lead to improvement of the discourse quality of their RA introductions and to better learning of rhetorical moves. Also, learning potential is reinforced when learners practice writing targeted discourse elements by

repeatedly modifying their output and when they engage in modified interaction with the program's available Help Options.

The next chapter continues the evaluation of IADE by examining its other three CALL qualities: Meaning Focus, Learner Fit, and Impact. The evaluation is also based on integrating multiple quantitative and qualitative data. First, findings about the program's ability to draw learners' focus to the functional meaning of their discourse are introduced. Then, it is revealed whether the types of automated feedback and help offered by IADE are fit for learners in the targeted instructional context. Last, ways in which interaction with IADE can impact the learners are discussed.

Chapter 6. Results on Meaning Focus, Learner Fit, and Impact of IADE

6.1. Introduction

This chapter covers the results regarding the strength of IADE's three CALL qualities: Meaning Focus, Learner Fit, and Impact. Each quality is discussed based on evidence gathered from quantitative and qualitative data: 88 responses to Likert-scale, questions, 83 responses to yes/no and open-ended survey questions as well as introspective data collected from 16 participants, each of whom provided one think-aloud protocol, one computer screen recording, one observation, and one semi-structured interview. The findings are detailed in view of the research questions posed:

- *Meaning focus* – What evidence suggests that the program's feedback directs students' attention towards the meaning of the targeted discourse forms?
- *Learner fit* – What evidence suggests that the formative feedback on targeted discourse forms is appropriate for students with the characteristics of the intended learners?
- *Impact* – What evidence suggests that students have a positive learning experience with IADE?

First, the results on Meaning Focus, or learners' focus on the functional meaning of the discourse, are presented. The evidence points not only to the strength of this quality of IADE, but also to its origin - the color-coded feedback. This type of feedback enhanced learners' noticing of negative evidence in the form of a mismatch between intended meaning and the colors in the feedback, which as a result directed their focus towards functional meaning. Such focus helped the participants identify a connection between functional meaning and move-specific phraseology and better construct new meaning.

Then, Learner Fit, or the appropriateness of IADE and its automated feedback for targeted learners, is examined. The findings show that the participants enjoyed working with IADE. Its feedback was perceived as appropriate, comprehensive, and facilitative. Such feedback characteristics as color-coded, iterative, individual-specific, and output-focused were viewed as very helpful. The data also suggests that the programs' Help Options increased its Learner Fit quality by reducing the level of task difficulty and that its method of automated evaluation encouraged learner autonomy.

Last, evidence of Impact, or the influence that IADE and its feedback had on learners, is discussed. Although some results uncovered negative aftermath, the participants perceived impact as mostly positive, acknowledging such beneficial effects as motivation, positive emotions, and development of new revision strategies. Alterations in their revision process also became transparent, finding explanations in the effects of IADE's two different types of feedback – numerical and color-coded, which affected the learners differently at cognitive, pragmatic, affective, and intrinsic levels.

6.2. Meaning Focus

The Meaning Focus (MF) quality of IADE and its automated feedback was investigated in order to find whether the feedback can draw learners' attention to the functional meaning of the research article Introduction discourse, which is realized through the functions of moves and steps. Evidence was sought in participants' responses to 3 Likert-scale questions, 3 yes/no and open-ended survey questions as well as in 16 think-aloud protocol/Camtasia recordings, 16 observations, and 16 semi-structured interviews. Table 6.1 summarizes the overall findings.

Table 6.1. Overall evidence of Meaning Focus

Data source	*N of participants*	*Data*	*Evidence for MF*	*No evidence for MF*
		Responses to:		
Survey data: Likert-scale q-ns	88	Q-n 8: When you were revising, did you think about the meaning you wanted to express?	98.9%	1.1%
		Q-n 9: When you were revising, did you notice that you meant one move, but the feedback showed you the color of a different move?	100%	0%
		Q-n 10: Did the feedback help you think of what you meant and better express your ideas?	100%	0%
Survey data: Yes/No & open-ended q-ns	83	Q-n 13: When you were trying to revise the moves in your introduction, did you think of the meaning you wanted to express in those moves? Why?	91.57%	8.43%
		Q-n 14: When you were trying to revise the moves in your introduction, did you notice that what you meant in your move/s was reflected by a different color in the feedback? If so, what did you think and what did you do?	100%	0%
		Q-n 15: Do you think the feedback helped better express the meaning you had in mind? How?	75.9%	24.1%
		Transcripts of:		
Introspective data: Think-aloud protocols/Camtasia	16	16 Think-aloud/Camtasia recordings	252/1227 idea units	
Semi-structured interviews	16	16 interviews	54/233 idea units	
Observational data: Observations	16	16 observations	77/460 idea units	

6.2.1. Learner perceptions on Meaning Focus

According to participants' responses to Likert-scale question 8, a total of 91.9% (87 students) focused on the meaning they intended to express. Only 1.14% (1 student) responded "not at all", while 80.68% (71 students) focused on meaning "a lot", 14.77% (13 students) – "somewhat", and 3.41% (3 students) – "a little" (see Figure 6.1).

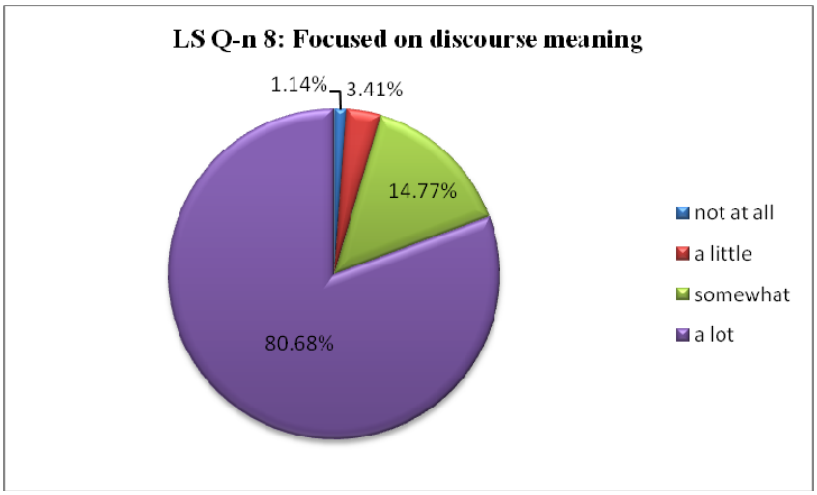


Figure 6.1. Results from Likert-scale question 8

Participants' answers to survey question 13, which inquired about their focus on the functional meaning of the rhetorical moves, were also more positive than negative. Of 83 respondents, 91.57% (76 students) focused on the functions of the moves, and 8.43% (7 students) did not think they did, explaining that their main goal was improving their draft and not thinking of the moves. For example, Student 7 wrote "Do you think every author in the world thinks about moves when he or she starts to write a research article draft?" Explanations of why the participants thought they focused on the meaning of the moves fell into several thematic categories. In their positive responses, the majority of the participants (68.42% or 52 students) reasoned that they thought of the functions of their sentences when

the color-coded feedback displayed a move-color differently than expected. Other students (11.84% or 9 students) explained that they had to think about the meaning of their moves in order to ensure the effectiveness of the communicative purpose of their discourse. In addition, 9.21% (7 students) realized that only by paying more attention to the functional meaning of the moves and steps they could better understand what they had been taught in class and therefore make better corrections to their drafts. No follow-up explanations were by provided by 10.53% (8 students). These findings are summarized in Table 6.2, which is followed by examples extracted from the data.

Table 6.2. Results from open-ended survey question 13

Reasons for focus on discourse form	*Participants*
Color-coded negative feedback	68.42%
Communicative purpose	11.84%
Enhanced understanding	9.21%
No reason	10.53%

- Color-coded negative feedback

  This was the hard part. Because, I cannot express well what I want to express, sometime the program shows other moves instead of what I want to express. And then I think what I'm saying (Student 42, survey question 13)

  I placed the meaning what I want to express first when some of moves were not in good consistent with the color of the results. (Student 52, survey question 13)

  I did, but I thought more when I saw the colored feedback. (Student 71, survey question 13)

  Before I revised I thought I always had the meaning in mind, but I didn't think about function. The meaning of words may be fine, but it's not enough for my message. When I see blue instead of red, I figure I don't have the right function. (Student 52, survey question 13)

- Communicative purpose

  If I want reader to know that I point to gap, I think about how I should say to point to gap. (Student 14, survey question 13)

In class I learned that introductions have many functions and I think about this because I want to make sure my introduction is not bad or misunderstand. (Student 19, survey question 13)

Of course I think about my meaning of moves because it's important for me to establish territory and gap and occupy gap very well in my introduction (Student 37, survey question 13)

They won't accept my paper if they don't understand where's centrality, where problem, where purpose for my research, so I think a lot to do this right (Student 81, survey question 13)

- Enhanced understanding

I only made good revise when I begin to think what I do with my sentence. (Student 53, survey question 13)

Actually, to know the moves you have to understand how they work, not only know them from lessons. That's why I thought of move meaning. (Student 7, survey question 13)

I thought meaning because my changes didn't work when I didn't think. Like if I just changed place of sentences, then the colors for sentence didn't change. (Student 21, survey question 13)

The meaning of moves is in definitions and I looked at that and tried to do what it say, so it makes sense now for me. (Student 81, survey question 13)

The role of the color-coded negative feedback became more evident due to the participants' responses to Likert-scale question 9. In their answers to this question, all the students indicated that they noticed having miscommunicated functional meaning due to a mismatch between intended and conveyed meaning displayed by IADE's colors. As shown in Figure 6.2, 44.32% (39 students) noticed such a mismatch "a lot", 40.91% (36 students) – "somewhat", and 13.64% (12 students) – "a little".

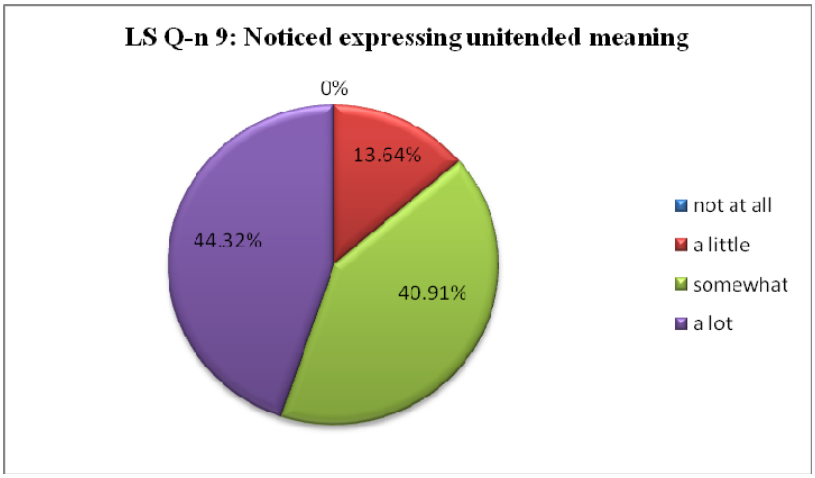


Figure 6.2. Results from Likert-scale question 9.

Survey question 14 yielded similar findings in that 91.57% (76 students) noticed that IADE's color-coded feedback displayed their moves in a color different than what they had in mind, and 8.43% (7 students) reported that this occurred, but not often. The participants explained how they reacted to such feedback. They commented on their thoughts (39.76% or 33 students) and actions (60.24% or 50 students) at the moment. Their thoughts included self-questioning as to what might have caused the mismatch (12.05% or 10 students), self-verification as to whether IADE is wrong in identifying the move and they are right in expressing its function (22.89% or 19 students), and self-planning speculations as to what should or should not be done when developing a given move (4.82% or 4 students). Participants' actions upon receiving color-coded feedback that was in disagreement with their communicative intent consisted of immediate attempts to modify their output (19.28% or 16 students), consulting IADE's Help Options (13.25% or 11 students), and searching for and using move-specific phraseology (27.71% or 23 students), which in fact also resulted in output modification. See table 6.3 and consider a few examples following it.

Table 6.3. Results from open-ended survey question 14

Reaction to negative color-coded feedback	*Participants*
Thoughts	39.76%
Self-questioning	12.05%
Self-verification	22.89%
Self-planning	4.82%
Actions	60.24%
Immediate modification of output	19.28%
Consulting IADE's HO	13.25%
Searching for and using move-specific phraseology	27.71%

- Self-questioning

  I only made good revise when I begin to think what I do with my sentence. (Student 3, survey question 14)

  I think why it reflected in a different color and how could I revise it in my desired color. (Student 50, survey question 14)

  I will think what is the problem with my sentence. (Student 65, survey question 14)

- Self-verification

  I think my own opinion again, and then revise it. (Student 11, survey question 14)

  I will check my sentences and confirm which move it should belong to. (Student 55, survey question 14)

  I will check the sentence and make sure that if I am right. If I am right, I just leave alone. Otherwise, I will change the sentence. (Student 74, survey question 14)

- Self-planning

  I think I need to use specific word to stress the move. (Student 64, survey question 14)

  I think that some sentences of article should not be in the part of moves like I have them. (Student 70, survey question 14)

  Like when I saw red in the beginning, I think you should have move 2 to start. You should have move 1 and I think how move 1 should be. (Student 6, survey question 14)

- Immediate modification of output
  I try to rewrite the sentence to what I really wanted it to be and in the same time balance the color (Student 38, survey question 14)

  I revised it until the color changed. (Student 63, survey question 14)

  I rewrite it and check it again. (Student 41, survey question 14)

- Consulting IADE's HO

  I tried to find out why there are differences and went back to the examples. (Student 57, survey question 14)

  I think there is something wrong. I read the papers in the LADE to figure out. (Student 76, survey question 14)

  I use "help options" to help me to changes the moves. (Student 25, survey question 14)

- Searching for and using move-specific phraseology

  I try to use some representative words to revise it. (Student 32, survey question 14)

  I will try to find out the differences between the moves and then use right expression on the move. (Student 30, survey question 14)

  Change the way to express it, especially using certain word to identify them. (Student 42, survey question 14)

In addition, it seemed that in their attempts to modify their output the learners were discovering a connection between certain vocabulary and the functional meaning of moves and steps, and therefore directed their attention to key words and expressions indicative of such meaning. For instance, Student 8 wrote "I realized that one word could change my thoughts. So, as I want to maintain my ideas of movements, I changed some expression to convert the sentences into other move or step." Student 12 implied a similar idea saying, "I was trying to change and insert specific words because I realized that my steps depends on

the right ‘words’”. Student 4 explained, “I was trying to find that if I used the wrong key word so it created the different meaning with what I really wanted to mean.”

In their answers to Likert-scale question 10, all 88 respondents reported that the feedback helped them focus on discourse meaning to a certain extent: 29.55% (26 students) thought it helped “a lot”, 54.54% (48 students) thought it “somewhat” helped, and 15.91% (14 students) thought it helped “a little” (see Figure 6.3).

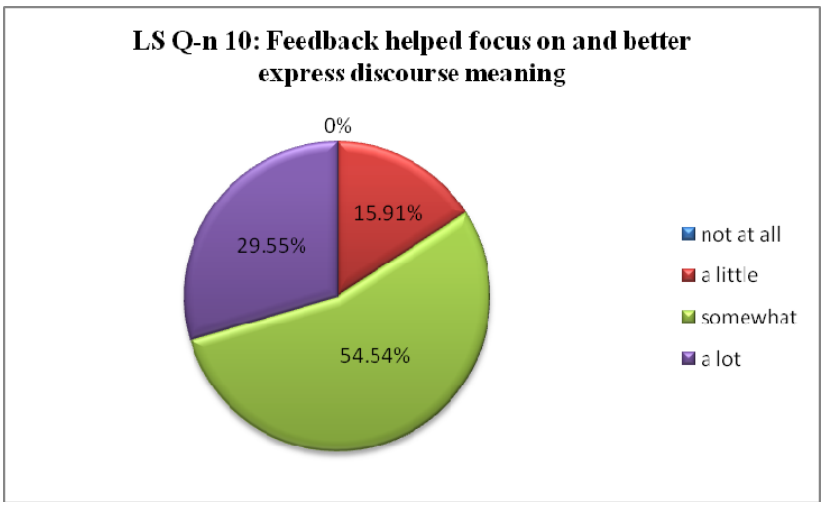


Figure 6.3. Results from Likert-scale question 10

In survey question 15, participants reflected on whether the feedback helped them better express the intended functional meaning. “Yes” answers were provided by 75.9% (63 students), “No” by 18.07% (15 students), and 6.02% (5 students) were not sure. Interestingly, negative answers were accompanied by such explanations as “It help me think about what my move actually is, but it doesn’t help much to fix it” (Student 57) or “If I see blue and not red, I know where I’m wrong, so that really helps (Student 13). It is not help to better express meaning because it’s not give suggestion how to express meaning better” (Student 64). In other words, the feedback was not entirely helpful to some students in this respect because it

only pointed to miscommunicated functional meaning, without providing a specific direction for remediation. On the other hand, the color-coded feedback appeared to have helped the majority of respondents better express the functional meaning of moves and steps for a number of reasons, which are summarized in Table 6.4. It made them think about what they were trying to express with their moves (12.7% or 8 students). It made them focus on negative evidence by displaying their moves in colors and thus making them aware of an existing mismatch between intended and expressed functional meaning (30.16% or 19 students). And it helped them identify a connection between functional meaning and lexical means of expressing it (57.14% or 36 students), an idea that supports the inference made earlier based on some of their responses to question 14.

Table 6.4. Results from open-ended survey question 15

Color-coded feedback	*Participants*
Helpful	75.9%
Focus on functional meaning	12.7%
Focus on negative evidence (mismatch in colors)	30.16%
Connection between functional meaning and move- specific phraseology	57.14%
Not helpful	18.07%
Not sure	6.02%

- Focus on functional meaning

  because I'm not sure about my own thoughts and colors make me think what I say. (Student 68, survey question 15)

  I could know the concept of the meaning I have in my mind. (Student 48, survey question 15)

  It helped me to pay more attention to how to say the moves. (Student 27, survey question 15)

- Focus on negative evidence

  The feedback tells me the moves sometimes different from what I try to express. (Student 44, survey question 15)

  It tells me whether I have expressed some sentences in a wrong way. (Student 21, survey question 15)

  I would find the improper explanation in my introduction. (Student 72, survey question 15)

- Connection between functional meaning and move-specific phraseology

  Especially with the appropriate words I should use to give determined meaning. (Student 61, survey question 15)

  I begin to focus on the key word and expression way. (Student 82, survey question 15)

  I can use more specific words to express the meaning in my mind. (Student 56, survey question 15)

6.2.2. Introspective data on Meaning Focus

The Meaning Focus quality of IADE and its automated feedback was also examined by analyzing the introspective accounts of learners' experience as reflected in the transcripts of 16 think-aloud/Camtasia protocols, semi-structured interviews, and observations. Of 1,227 idea units identified in the think-aloud/Camtasia data, 252 idea units were coded for Meaning Focus, which averages 15.75 idea units per participant. The observation transcripts contained 77 Meaning Focus idea units of a total of 460 idea units identified in this data, averaging to 4.8 idea units per student. The interview transcripts had 54 Meaning Focus idea units out of 233, with an average of 3.4 per student.

All these data sources contained evidence of phenomena that were mentioned by the participants in their survey responses. Judging by the 16 students' utterances and actions reflected in the recordings, IADE's color-coded feedback played an important role in

directing the learners' focus to the meaning of their discourse. Moreover, the analysis of the MF idea units indicated that the learners' focus on functional meaning was manifested in a number of ways. Having seen the colors marking their moves, the participants took time to reflect about how effectively or ineffectively they expressed the intended meaning. They also focused on meaning by noticing that in some cases the intended functional meaning was displayed with the color of a different move. In addition, what was indicative of meaning focus is that the learners seemed to realize that the effectiveness of expressing functional meaning is directly related to certain lexical choices and, therefore, to further modify their output in view of the move-specific phraseology that they could find in IADE's Help Options. Percentages for each of these themes per data source are given in Table 6.5.

Table 6.5. Themes indicative of Meaning Focus identified in think-aloud/Camtasia, observations, and interviews transcripts

Themes that emerged	*TA/Camtasia*	*Observations*	*Interviews*
Noticing a mismatch between intended and expressed functional meaning	29.76% (75 idea units)	32.47% (25 idea units)	44.45% (24 idea units)
Reflection on functional meaning	17.46% (44 idea units)	40.26% (31 idea units)	22.22% (12 idea units)
Connection between functional meaning and lexical choice	37.7% (95 idea units)	16.88% (13 idea units)	33.33% (18 idea units)
Construction of new functional meaning	15.08% (38 idea units)	10.39% (8 idea units)	
Total idea MF units	100% (252 idea units)	100% (77 idea units)	100% (54 idea units)
Total idea units	1227	460	233

Consider a few examples from the transcripts, which show how the students verbalized their thoughts when they noticed a mismatch, pondered over the functional meaning of their

moves, connected it with certain vocabulary, and used it to construct better functional meaning.

- Noticing a mismatch bw intended and expressed functional meaning

  The last sentence is recognized as m1, but I meant it as m3. (Student 27, think-aloud protocol and Camtasia transcript)

  What's the matter? [highlights and reads the blue text meant as m3] I think this sentence at least description. (Student 40, think-aloud protocol and Camtasia transcript)

  For the last two sentences, I think they are belong to m3 as well. And they were classified as m2. (Student 64, think-aloud protocol and Camtasia transcript)

  Begins reading and thinking about the first sentence and notices that she meant m2. (Student 28, observations transcript)

  Isn't sure whether a sentence is m3 or not. Looks up definitions and understands that it's actually m2, not m3. (Student 53, observations transcript)

  I think I revised more when I see that color is wrong. Like, like…I see green in first sentence and I know it's wrong because I didn't mean start with move 3. (Student 58, interview transcript)

  Why I comment every sentence? Because I want to find where my moves are different color. (Student 62, interview transcript)

- Reflection on functional meaning

  Yeah, in this sentence I mean to say why our method is helpful for such applications because the MRI image is time-consuming and expensive to take completely samples. So we use this method to reduce the samples we need to take. So I mean to demostrate that this specific topic, in other words, this method is helpful to this. So, I use m2 (Student 30, think-aloud protocol and Camtasia transcript)

  Yeah, in this sentence I mean to say why our method is helpful for such applications because the MRI image is time-consuming and expensive to take completely samples. So we use this method to reduce the samples we need to take. So I mean to demonstrate that this specific topic, in other words, this method is helpful to this. So, I use m2 (Student 32, think-aloud protocol and Camtasia transcript)

This is... I think this is m2 ... [cursor over this sentence, which is red] this is m2. I think it's right because problem ... current... this is analyze the general... explains the ... this is a problem. (Student 54, think-aloud protocol and Camtasia transcript)

Justifies the use of m3 in the middle of the text, explaining what the intended effect is. (Student 29, observations transcript)

Looks back at his sentence and decides that it sounds like m1 Topic Prominence (Centrality). (Student 58, observations transcript)

Goes on to the next sentence, reads it carefully, and agrees that it should be green because it shows the purpose of the article. (Student 64, interviews transcript)

Oh, that part I delete because it shows the advantage of other method, but I want just to show a step of the structure of the paper where my method is. (Student 40, interviews transcript)

- Connection between functional meaning and lexical choice

I'll see if there is some examples. [reads aloud from the examples found and emphasizes "we found that...", "in particular show that", "in our experiments"] ok, i think that should be easy. [speaks aloud and emphasizes] the contribution of our work, of our proposed system... (Student 43, think-aloud protocol and Camtasia transcript)

[opens definitions with the few examples provided there] the aim of this study was bla-bla... the purpose of the present study was... ok. (Student 61, think-aloud protocol and Camtasia transcript)

Let's see more example. [opens another annotated text and clicks on green sentences] definition m3. outlining structure... we have organized this paper. here... this paper... yeah. [checks the definitions in the HO] so, they have a... they have strong words to indicate that the research results belongs to this paper. (Student 58, think-aloud protocol and Camtasia transcript)

Thinks about using move-specific vocabulary ("What indicator words can I use?") (Student 27, observations transcript)

Wants to express m2 Hypothesis and changes "can" to "could". (Student 44, observations transcript)

I didn't really think you have to write steps with some key words. I figure out when I look again at example. (Student 62, interviews transcript)

When I saw another color, I change some words. You know, it doesn't work just with words, but … like… I changed one sentence with negative, and it show the problem better. It become red. (Student 64, interviews transcript)

- Construction of new functional meaning

Replaces "can" with "will", replaces "us" with "the industry". (Student 29, think-aloud protocol and Camtasia transcript)

[goes down to the revision box. highlights 'may also' and changes it to "may contribute to"] I changed just ... m2. I think it maybe will let the problem change it into m3. (Student 65, think-aloud protocol and Camtasia transcript)

I'm trying to use the word "assume" as an indicator of m2 hypothesis. (Student 27, think-aloud protocol and Camtasia transcript)

Adds words that better signal the moves. (Student 30, observations transcript)

Then he thinks that "this model" may be perceived as a continuation of the previous Review sentence and doesn't not clearly indicate that it's the model used in his study. Therefore, he changes "this" to "our". (Student 40, observations transcript)

The think-aloud/Camtasia and observations transcripts helped better understand the process of focus on functional meaning. Such focus typically originated with the color-coded feedback drawing the learners' attention to the mismatch between what they meant to do in their discourse and what they actually did. That triggered their reflection on intended meaning, which sometimes had not even been considered during the writing of the first draft. When trying to find a way to better express what they meant in their moves, the learners tended to look up examples in IADE's Help Options, where they noticed certain lexical items that were clearly indicative of certain functions, and then to use them in their own sentences. Generally, these lexical modifications lead to positive color-coded feedback, which displayed the color of the intended move. As a result, the learners appeared to have discovered a relation between certain vocabulary indicative of certain move functions. Also, employing

move-specific words and expressions helped them construct functional mening more effectively.

This phenomenon is exemplified in Figure 6.4, where Student 40 first sees that his move 3 has the color of move 1 and then reflects about the functional meanings of a number of rhetorical steps within move 1, move 2, and move 3 that he was trying to convey – establishing the field through a preview of previous research, establishing a niche by highlighting a problem, and occupying the niche by describing the model he had developed in his study. Then, viewing examples in the annotated corpus given in IADE's Help Options, he realizes that researchers refer to their own work with the help of personal pronouns and, therefore, uses "our" to begin describing his model.

Noticing a mismatch between intended and expressed functional meaning
So this part I want it to be m3, but uh it still displays in m1. I don't know why.

Reflection on functional meaning
See here, this part. [highlights a blue part of the text] in my mind I wanted to write about method. So I think I describe the method, but it's m1. I don't know why. So, see... [highlights a blue sentence] this sentence may be review to describe the reason, but in this model ... [highlights a blue sentence] this sentence I think I want to describe the model, how it works. So this, uh, some parameters and ... the method. Actually I don't know which move it will be. I just want to first ... in my mind, I think first I describe the field, uh, the technology in this field. Second, I do some reviews, and later I um... from the review I find some um... maybe not perfect in the previous research, so I can do something more. And in this part... so I describe what I want to do in this paper, and briefly, briefly introduce my method.

Connection between functional meaning and lexical choice
Let me see some examples [opens AC in the HO. looks at green sentences and their steps. highlights some.] Hmmm. Let me find some more examples to describe the model. [opens the AC, scrolls down to the green text in an annotated introduction, checks the steps by clicking, highlights parts of the green text. opens another annotated text and does the same] maybe I can write it as description. [speaks aloud and emphasizes] "our work", "of our proposed system"... Let's try this first.

Figure 6.4. Example of the process of focus on meaning (think-aloud protocol/Camtasia transcript, Student 40)

Construction of new functional meaning
[reads his text, then makes changes: "In this model, tungsten cones with different sizes are used to enhance the optical field." --> In our model, several tungsten cones with different sizes were chosen to enhance the optical field." [submits]

Figure 6.4. Example of the process of focus on meaning (think-aloud protocol/Camtasia transcript, Student 40) (continued)

While this sequence appeared to be a prominent part of learners' revision process, it did not always occur in a linear fashion. What seemed to be more consistent was the causative effect of two elements: noticing a mismatch between intended and expressed meaning lead to reflection on functional meaning, and making connections between functional meaning and lexical choice lead to construction of new discourse meaning. Consider the following excerpts:

- Noticing a mismatch between intended and expressed functional meaning →

  → Reflection on functional meaning

  And here I find another misunderstanding. → [reads the sentence] My intent is to ... I guess it could be m2 or m3, but not m1. (Student 27, think-aloud protocol and Camtasia transcript)

  Hm… again blue. → I try to make it like m3, but it really sounds generalized things. (Student 65, think-aloud protocol and Camtasia transcript)

  Understands that the sentence she thought should be m3 is not really m3. → Then she thinks and decides that it's probably Justification m2. (Student 28, observations transcript)

  He still has a problem. Part of m3 appears as m1. → He thinks that maybe it sounds as Review (which it does). (Student 40, observations transcript)

  When the color is not right, → I think what step I want it in that sentence. (Student 58, interviews transcript)

Yeah, I know, so many times I try to fix it! → It's hard to figuring out if I want value but I think software understands something in move 1. (Student 29, interviews transcript)

- Connection between functional meaning and lexical choice →

  → Construction of new functional meaning

  Now I found that I add one sentence for m2, but without the "although" my previous sentence changed to m1. → I'll add "although" because I want to show contrast. (Student 32, think-aloud protocol and Camtasia transcript)

  OK, it is important to address the problem… "drawback" is one of those words that point to something is problem. → [adds "drawback]. (Student 43, think-aloud protocol and Camtasia transcript)

  Wants to express m2 Hypothesis → and changes "can" to "could". (Student 44, observations transcript)

  Goes to Definitions in m3. Utters all the key words in the given examples and realizes that there are words that indicate m3. → Adds "in this paper" and re-submits. (Student 61, observations transcript)

It was also interesting to note that once the learners realized that vocabulary carries functions, they began paying more attention to their use of vocabulary. For instance, researcher's observation notes for Student 30 read, "He spots the word 'previous', thinking that it is the reason for why his intended m3 appears as m1. Hypothesizes about what would be a good lexical replacement that would indicate the intended step in m3." For Student 29, the researcher wrote, "Thinks that 'can' makes m3 Value sound like Hypothesizing. Wants to try 'contribution is'." Here are similar examples from the other two data sources:

I want to change... this move to be m1. I think 'however' maybe … because this word is included in m2. (Student 65, think-aloud protocol and Camtasia transcript)

So, I'm gonna uh revise the definition of the moves in order to ... to search new words to to to change from m1 to m2. (Student 62, think-aloud protocol and Camtasia transcript)

I think the last two sentences is m3, but the program marked the sentences as blue. [highlights a blue sentence] this sentence should be m3. hmmm... There's something wrong with my ... uh... language? I think it's because I use "those"... maybe I need "these" to say it's results in this study? (Student 61, think-aloud protocol and Camtasia transcript)

For example, to fix the color one time I thinking maybe I should use "in our study" (Student 30, interviews transcript)

I start thinking what key word should I use? (Student 27, interview transcript)

The semi-structured interviews provided additional insights into learners' focus on discourse meaning. One such insight is that the students did not think of meaning when they started revising their draft; they seemed to focus more consistently on functional meaning later in the revision process, once they realized the importance of functional meaning after having come across discrepancies displayed by the color-coded feedback. When asked why he referred to specific steps more frequently towards the end of the think-aloud, Student 43 said, "I didn't even think about moves and steps when I first write my introduction. But when I see no red, it was like something is missing, so I start thinking what. I don't know, so I look at definitions and examples. So I realize it's not so simple. If I don't think about moves or steps, it affect the whole thing."

Another observation is that, initially, noticing a mismatch in the colors was mostly accidental, and, as the revision continued, it became intentional. In other words, the learners began to intentionally verify their sentences to see if they are displayed in "the right color". For example, Student 29 explained, "I don't even think much about colors at beginning. I looked at percent for my field. When I want to add to my move 2 to make it increase, I saw that one sentence I think is move 2 was blue. Then I start checking every time because if it's wrong color, it not do what is supposed to do."

A less encouraging insight is related to discovering the lexical realizations of move/step functions. Nine out of 16 students mentioned that once they realized that certain words can help them build certain moves, they tended to rely more on making lexical modifications. Along these lines, Student 64 said, “Yes, I used those words there because they work. Actually, this is good that I know that because after that I changed many words many times.” Indeed, having this awareness is a positive thing; however, learners’ tendency to replace some vocabulary items with others, making this their main revision strategy, is limiting and undesirable.

6.2.3. Overall evaluation of Meaning Focus

Overall, the findings indicate that IADE and its automated feedback possess a strong Meaning Focus quality, judging by the evidence summarized in the figures below. Like for Language Learning Potential, participants’ scale answers were converted to evaluative comments - “not at all” was equaled to poor evidence, “a little” – to weak, “much” – to good, and “a lot/very much” – to excellent. Figure 6.5a shows that a total of 52% of participants perceived the Meaning Focus quality of IADE as “excellent”, and 37% as “good”. Similarly, Figure 6.5b shows that the evidence that the learners focused on discourse meaning, perceived the feedback as being helpful in this respect, and noticed having expressed unintended meaning is mainly “good” and “excellent”.

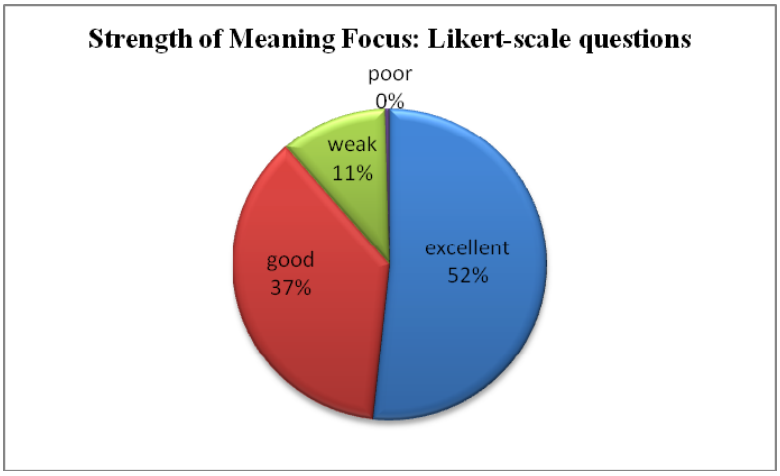


Figure 6.5a. Strength of Meaning Focus in Likert-scale data (N=88)

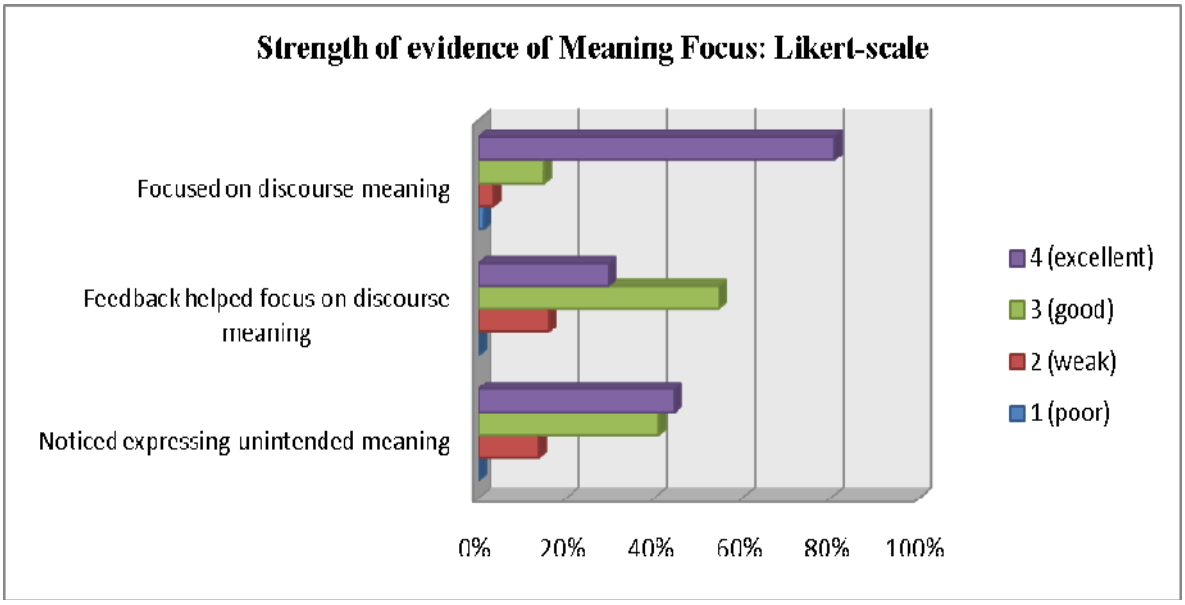


Figure 6.5b. Strength of Meaning Focus evidence in Likert-scale data (N=88)

In yes-no survey questions, cumulative evidence of Meaning Focus amounted to 89.16% (see Figure 6.5c). In addition, the think-aloud/Camtasia, observations, and interview transcripts all contained idea units that shed light on how and why learners focused on discourse functional meaning (see Figure 6.5d).

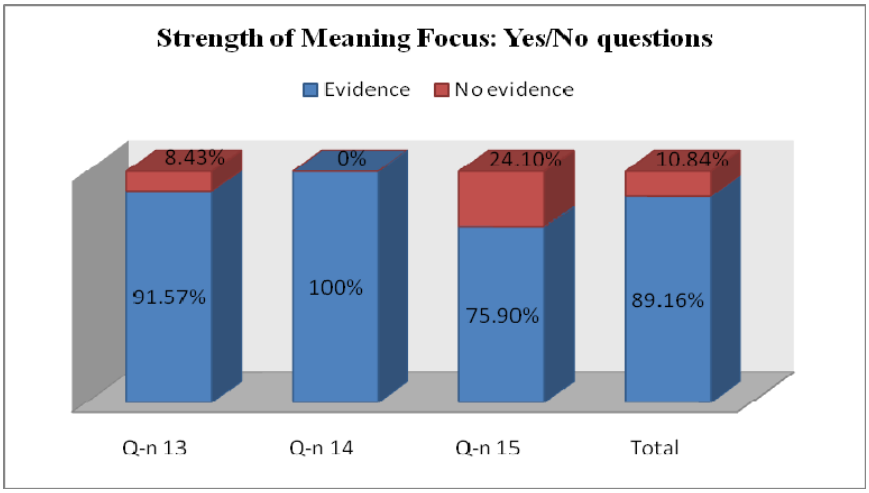


Figure 6.5c. Strength of Meaning Focus in Yes/No survey data (N=83)

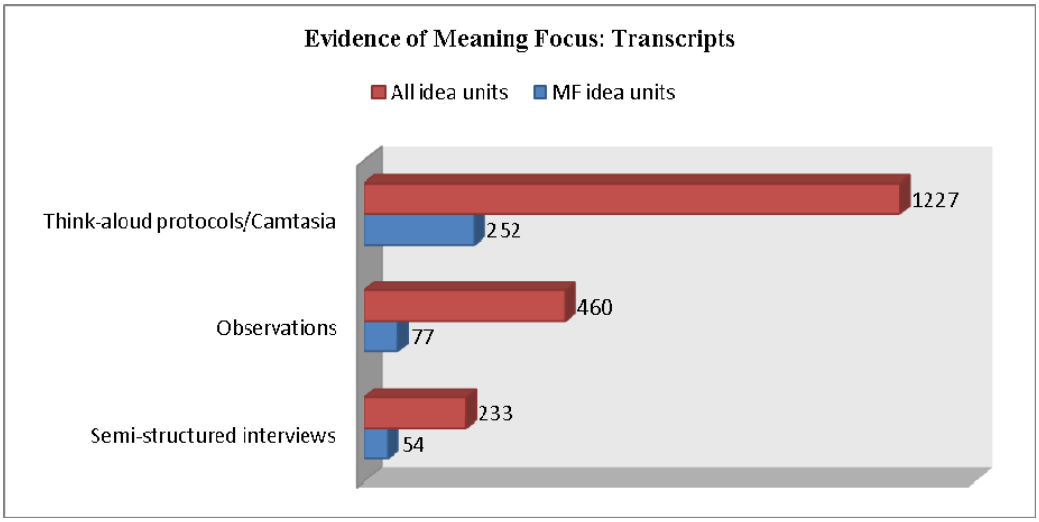


Figure 6.5d. Evidence of Meaning Focus in transcripts data (N=16)

The results in the form of themes that emerged in different qualitative data indicate that it was the color-coded feedback that made IADE's Meaning Focus strong. In particular, the transcripts showed that the learners noticed having expressed unitended meaning, which is supported by the noticing a mismatch theme in open-ended survey questions. According to question 14 responses, one way in which the learners reacted to the negative color-coded feedback was thinking about the cause of the mismatch and about the intended function,

which is reinforced by the reflection on functional meaning theme in the transcripts. Another way they behaved upon the reception of such feedback was taking some actions such as searching for move-specific phraseology in IADE's Help Options and modifying their output, which is exactly what was observed in the transcripts. To conclude, the Meaning Focus results are helpful in gaining a better understanding of why and how such focus occurred.

6.3. Learner Fit

Another CALL quality investigated was Learner Fit (LF), i.e. the appropriateness of IADE and its automated feedback for learners with targeted characteristics in NNS graduate-level academic writing contexts. Data analyzed for this purpose consisted of 5 Likert-scale questions and 5 yes/no and open-ended survey questions, which reflected the perceptions of all the respondents. Additional introspective insights were gained from 16 think-aloud protocol/Camtasia recordings, 16 observations, and 16 semi-structured interviews. As presented in Table 6.3, all the sources yielded positive evidence of Learner Fit.

Table 6.6. Overall evidence of Learner Fit

Data source	*N of participants*	*Data*	*Evidence*[5] *for LF*	*No evidence for LF*
Survey data: Likert-scale q-ns	88	Responses to:		
		Q-n 11: Did you understand the feedback provided by the program?	100%	0%
		Q-n 12: How helpful did you find the characteristics of the feedback?	99.5%	0.5%

[5] For open-ended survey questions 19 and 20, Learner Fit evidence represent the percentages of answers that indicated complete satisfaction. No evidence is represented by percentages of responses with criticisms and desired characteristics of IADE's feedback. Percentages of responses that had no opinion are not included.

Table 6.6. Overall evidence of Learner Fit (continued)

Data source	*N of participants*	*Data*	*Evidence for LF*	*No evidence for LF*
		Responses to:		
Survey data: Likert-scale q-ns	88	Q-n 13: Was the feedback on discourse moves appropriate for you?	100%	0%
		Q-n 14: Did the program motivate you to improve your introduction draft?	100%	0%
		Q-n 16: Did you like using IADE?	100%	0%
Survey data: Yes/No & open-ended survey q-ns	83	Q-n 16: Was the feedback appropriate for you? Why?	90.36%	9.64%
		Q-n 17: Was individualized feedback based on a comparison with the norms of writing in your field, motivating for you? Why?	80.72%	19.28%
		Q-n 18: What did you like about the feedback the most?	100%	0%
		Q-n 19: What did you like about the feedback the least?	32.53%	57.83%
		Q-n 20: What other kinds of feedback would be helpful to you?	37.5%	51.8%
		Transcripts of:		
Introspective data: Think-aloud protocols/Camtasia recordings	16	16 Think-aloud/Camtasia recordings	121/1227 idea units	
Semi-structured interviews	16	16 interviews	43/233 idea units	
Observational data: Observations	16	16 observations	84/460 idea units	

6.3.1. Learner perceptions on Learner Fit

The survey questions targeting IADE's Learner Fit quality revealed that the program and its feedback were perceived by learners as appropriate for a number of reasons. First of

all, according to the answers to Likert-scale question 13 and to yes/no survey question 16, most of the respondents thought that the feedback in general was appropriate for them. In the former question, none of the respondents chose “not at all”; 27.27% (24 students) chose “very” appropriate, 59.09% (52 students) chose “somewhat”, and 13.64% (12 students) chose “a little” (see Figure 6.6).

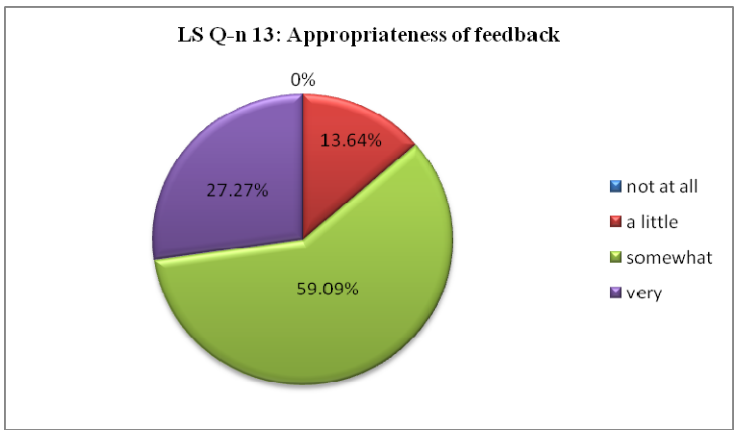


Figure 6.6. Results from Likert-scale question 13

For question 16, 90.36% (75 students) reported that the feedback was appropriate, 6.02% (5 students) thought the opposite, and 3.61% (3 students) were not sure. Those who did not find the feedback appropriate explained that “it gives too little information” (Student 38), that it was “not very specific” (Student 35), or that it was “not really helpful because the ratios of different moves vary in different journals” (Student 51). Except for 7 students, the positive answers were also accompanied by explanations, which mostly pointed to certain reasons why the students thought that IADE’s feedback was appropriate. These reasons are listed in Table 6.7. The most common theme, mentioned by 44% or 33 students, was the individual orientation of the feedback, i.e. the fact that the feedback was provided individually for learners and that it was based on their individual academic discipline.

Another 28% (21 students) appreciated that both the colors and the percentages helped them see imperfections in their writing. For 18.67% (14 students) the feedback was appropriate because it was clear.

Table 6.7. Results from open-ended survey question 16

Reasons for perceptions of feedback appropriateness	*Participants*
Display of negative evidence in feedback	28%
Clarity of feedback	18.67%
Individual orientation	44%
No reason	9.33%

- Display of negative evidence in feedback

  It was appropriate for me because I have lots of sentences reflecting the wrong color. (Student 50, survey question 16)

  It's appropriate for me because it shows my problem. (Student 44, survey question 16)

  because it grabbed the major issues. (Student 54, survey question 16)

- Clarity of feedback

  It is good. Because it is easy explanation. (Student 67, survey question 16)

  it was clear to me. (Student 28 survey question 16)

  There are two different feedbacks based on the moves, and both to the point and both easy for me to understand. (Student 80, survey question 16)

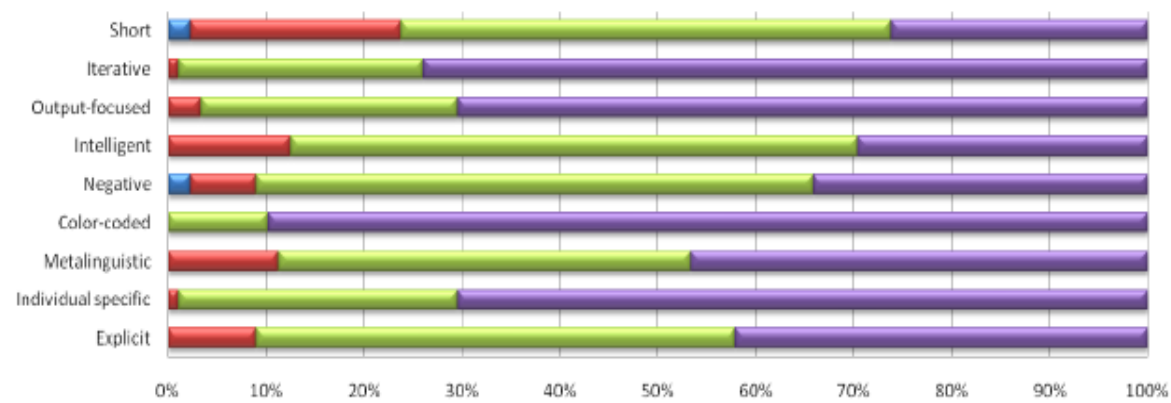
- Individual orientation

  because it told me about my writing and writing in my field (Student 83, survey question 16)

  Because I really want to know how is my steps organized in each move. (Student 27, survey question 16)

  because it gave me some regular situation in my field for the introduction. (Student 4, survey question 16)

In Likert-scale question 12, the respondents evaluated the helpfulness of nine characteristics of IADE's feedback. Figure 6.7 shows that the color-coded feedback was rated as "most helpful" by 89.77% (79 students) and as "somewhat helpful" by 10.23% (9 students). Another three feedback characteristics rated relatively high were: iterativity, with 73.86% of "very helpful" responses and 25% of "somewhat helpful"; focus on learner's output, with 70.45% "very helpful" and 25% "somewhat helpful"; and individual specificity, with 70.45% "very helpful" and 28.41% "somewhat helpful". Other features, such as intelligent, negative, and metalinguistic were mostly perceived as "somewhat helpful". The conciseness and explicitness of the feedback seemed to be "of little help" to 21.59% (19 students) and to 9.09% (8 students), respectively. These two characteristics were of "no help" for two students. Similarly, 12.5% did not think that IADE's intelligence was of much help.



	Explicit	Individual specific	Metalinguistic	Color-coded	Negative	Intelligent	Output-focused	Iterative	Short
not helpful	0.00%	0.00%	0.00%	0.00%	2.27%	0.00%	0.00%	0.00%	2.27%
of little help	9.09%	1.14%	11.36%	0.00%	6.82%	12.50%	3.41%	1.14%	21.59%
somewhat helpful	48.86%	28.41%	42.05%	10.23%	56.82%	57.95%	26.14%	25.00%	50.00%
very helpful	42.05%	70.45%	46.59%	89.77%	34.09%	29.55%	70.45%	73.86%	26.14%

Figure 6.7. Results from Likert-scale question 12

The positive evaluation of the negative, explicit, and individual-specific characteristics of IADE's feedback is consonant with the themes that emerged in participants' answers to open-ended question 13, which reinforces this piece of evidence on feedback appropriateness. In addition, the form in which the feedback was provided, i.e. color-coded and numerical, allowed most of the participants to understand it well: 52.27% understood the feedback "completely", 40.91% - "somewhat", and 6.82% - "a little" (see Figure 6.8).

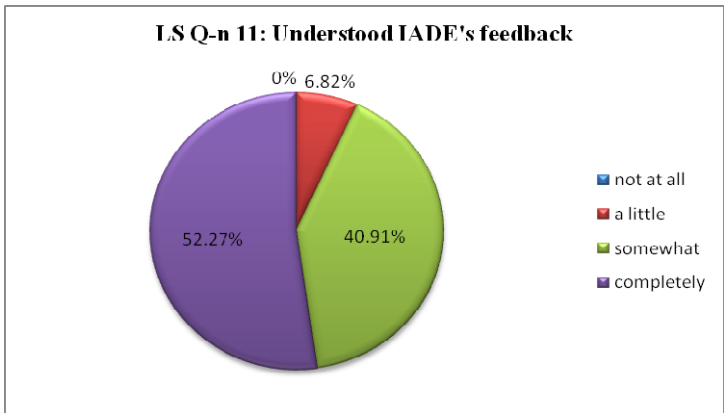


Figure 6.8. Results from Likert-scale question 11

Survey question 17 asked the learners whether the fact that the feedback was based on a comparison with their discipline motivated them to revise their draft. Most of the respondents (80.72% or 67 students) answered "yes", 15.66% (13 students) answered "no", and 3.61% (3 students) were not sure. The "why" follow-up question was addressed only by 59.04% (49 students). The percentage of "no" responses can be explained by the answers of 16.33% (8 students), which clarified that comparison with disciplinary norms in IADE's feedback was not a major motivational factor because their particular discipline (e.g. Bionanotechnology, Ergonomics, Materials Science, etc.) was not represented in IADE, or

because the participants believed that the comparison couldn't have captured all the peculiarities in different journals.

- Discipline not represented in IADE

  My field is not in the list. (Student 68, survey question 17)

  I cannot answer for it. Actually there is not field I am studying for. I used just engineering, but it's too many kinds of engineering. I just focus most on colors for moves. (Student 12, survey question 16)

  My field is actually located in between math and computer. What field do I have to choose? There're so many sub fields inside of Computer Science, those cannot be classified as just computer science. Both Engineering and Science are big topics (Even biology, genetic, mathematics) of computer science. (Student 34, survey question 16)

- Different writing norms in different journals

  Not very much, because the ratios of different moves vary in different journals (Student 51, survey question 17)

  Not much, because it changes from journal to journal I think we cannot rely on that (Student 39, survey question 17)

  Not quite, math paper is different in different publications (Student 20, survey question 17)

The motivation of the other 83.67% (41 students) could be explained in view of two main reasons: the learners' intrinsic desire to follow disciplinary norms of their academic community and the extrinsic influence of the interaction with the intelligence of the program. In other words, they felt motivated because they wanted to write like the academia in their field or because they engaged in learner-computer interaction that approximated interaction with human professionals. For example:

- Desire to follow disciplinary norms

  because as an entrance level research, I would like to keep consistent with the norms. (Student 54, survey question 17)

  I like this because we me a goal to follow and be accord with my field. (Student 47, survey question 17)

  I think it is the best option to use comparison article on my own field because I want to write like my field. (Student 71, survey question 17)

- Interaction with the intelligence of the program

  It can instruct me individually, and adapt for me like a reviewer. (Student 55, survey question 17)

  It's like the professionals give me feedback. (Student 40, survey question 17)

  It made me know how to improve and want to improve because it knows my field and tells me right away correctly. (Student 83, survey question 17)

Another angle on Learner Fit was taken by the survey questions about learners' experience with IADE and its feedback, which asked whether they liked it or not and why. According to scale question 16, all the participants enjoyed working with IADE; 73.86% (65 students) liked it "very much" and 26.14% (23 students) "somewhat" liked it.

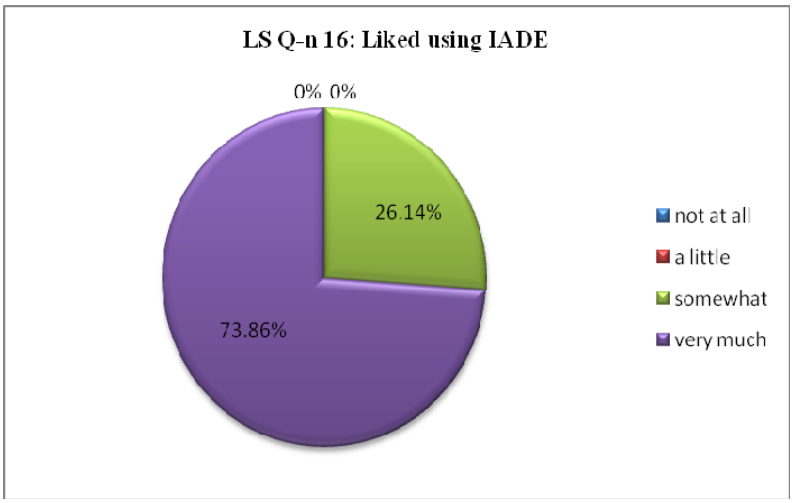


Figure 6.9. Results from Likert-scale question 16

Open-ended survey questions 18 and 19 revealed what the participants liked and what they disliked about IADE's feedback. As it can be seen in Table 6.8, the form in which the feedback was displayed was mentioned the most. The color codes were preferred by 33.74% (28 students), and the comparison of participants' writing to the norms in their discipline given in percentages and comments were mentioned by 36.15% (30 students). Another 9.64% (8 students) referred to both these characteristics of the feedback. In addition to these, a few participants pointed out that they liked that the feedback showed them where their discourse was deficient and that it was instant, individualized, intelligent, and clear. These answers, when juxtaposed with the answers scale to question 12 where the participants rated the helpfulness of feedback qualities, corroborate the findings indicating that IADE's color-coded feedback was appropriate and helpful for learners, as were its individual-specific and output-focused qualities expressed through the comparison with discipline and individualization themes identified in open-ended question 18. The negative, explicit, and intelligent features were also thought to enhance the Learner Fit aspect of IADE's feedback in both these data sources.

Table 6.8. Results from open-ended survey questions 18 and 19.

Likes		*Dislikes*	
Color-codes	33.74%	Degree of specificity	19.28%
Comparison with discipline	36.15%	Lack of feedback on steps	30.12%
Color-codes & comparison with discipline	9.64%	Accuracy of analysis	8.43%
		Nothing	32.53%
Negative evidence	6.02%	Not sure	9.64%
Immediacy	3.61%		
Individualization	3.61%		
Intelligence	4.82%		
Clarity	2.41%		

Table 6.8 also lists the qualities of IADE's feedback that the participants liked the least according to their responses to open-ended question 19. The fact that 32.53% (27 students) said that there was nothing they did not like can be considered as positive evidence of Learner Fit, and the comments on dislikes should be informative for IADE's further instructional implementation and design improvement. The feedback would be more appropriate for learners if it provided more details on their written performance, judging by the opinion of 19.28% (16 students), who thought that it was not specific enough, and of 30.12% (27 students), who expected to receive feedback at the level of steps, not only at the level of moves. A few students (8.43% or 8 students) questioned the accuracy of automated analysis, based on which the feedback was generated. Some had realized that the program relies on a lexical identification approach and viewed that as limiting. Participants' comments in this question help better understand why some of them evaluated the helpfulness of short, explicit, metalinguistic, and intelligent characteristics of IADE's feedback as either of little help or not helpful (see Figure 6.7). Following are examples of open-ended responses to questions 18 and 19.

- Likes

  I liked how the colors showed me what my writing means. (Student 5, survey question 18) → Color-codes

  the comparison of my introduction with corpus in my field. (Student 36, survey question 18) → Comparison with discipline

  Colors and percentages in my field. (Student 28, survey question 18) → Color-codes & comparison with discipline

  Tell me my shortcomings. (Student 58, survey question 18) → Negative evidence

  It's back in a few seconds. (Student 62, survey question 18) → Immediacy

It is for one person -me. (Student 51, survey question 18) → Individualization

I like everything. It's very smart. (Student 30, survey question 18) → Intelligence

It's simple and clear (Student 53, survey question 18) → Clarity

- Dislikes

I think it does not give me enough information. (Student 38, survey question 19) → Degree of specificity

I don't know about the steps (Student 45, survey question 19) → Lack of feedback for steps

Sometimes it recognizes the move incorrectly. (Student 35, survey question 19) → Accuracy of analysis

There is no inappropriate feedback. It's all good for me. (Student 5, survey question 19) → Nothing

I don't know really. (Student 5, survey question 19) → Not sure

Some aspects that the students disliked about IADE's feedback were also mentioned in their responses to open-ended question 20, where 32.53% (27 students) explained that they would have liked to receive feedback that evaluated both the moves and the steps. Another 12.05% (10 students) suggested that the feedback should be provided not only at the level of discourse, but also point to language errors such as grammar, word choice, and expression, and even spelling. For instance, Student 32 wrote that he would have liked "Feedback about grammar and words," Student 54 wanted the program to "provide grammar and spelling detection," and Student 40 wanted "to know if some words used are not professional." A few participants (4.82% or 4 students) thought that a specific remediation direction such as "maybe some tips of revising" (Student 59), "revising guidelines" (Student 63), or "further suggestion about how to improve the ratio of different moves" (Student 86) would enhance

the degree of feedback helpfulness. Perhaps, the Revision Tips in IADE's Help Options were not very helpful to these learners, or they may not have used those revision hints sufficiently since the records in IADE's database show that these particular students only accessed them one or two times. Two other participants wished they knew "How the program determines which move the sentence is" (Student 57) and "Why IADE classify a certain sentence to a move" (Student 37), although it is not clear whether they wanted this information as part of the feedback or whether they were just curious to know more about the program's intelligence. The remaining 37.35% (31 students) were satisfied with the feedback and had nothing to suggest, and 10.84% (9 students) did not know what would make the feedback better or more appropriate for them. Table 6.9 summarizes these findings.

Table 6.9. Results from open-ended survey question 20

Desired feedback qualities	*Participants*
Feedback on steps	32.53%
Feedback on language errors	12.05%
Specific direction for remediation	4.81%
Explanation of program intelligence	2.41%
None	37.35%
No opinion	10.84%

6.3.2. Introspective data on Learner Fit

Consistent with the qualitative analyses of IADE's other CALL qualities, evidence of Learner Fit was also sought in the transcripts of 16 think-aloud/Camtasia protocols, observations, and semi-structured interviews. The first of these three data sources contained 121 idea units coded for Learner Fit, averaging 7.6 idea units per student. The observations transcripts had 84 such units, and the interviews had 43, with an average of 5.3 and 2.7 units per student, respectively. These data provided insights that did not surface in the survey

questions, but present equal interest for the purpose of evaluating the appropriateness of IADE and its feedback for targeted learners. The themes that recurred in these three sources are presented in Table 6.10 and discussed below.

Table 6.10. Themes indicative of Learner Fit identified in think-aloud/Camtasia, observations, and interviews transcripts

Themes that emerged	*TA/Camtasia*	*Observations*	*Interviews*
Facilitating problem-solving	85.95% (104 idea units)	88.09% (74 idea units)	20.93% (9 idea units)
Expressed difficulty	5.78% (7 idea units)	8.33%% (7 idea units)	
Observed difficulty	80.17% (97 idea units)	79.76% (67 idea units)	
Learner control	14.05% (17 idea units)	11.91% (10 idea units)	13.95% (6 idea units)
Appropriateness of feedback for task completion			23.26% (10 idea units)
Suitability for individual learner characteristics			32.56% (14 idea units)
Appropriateness of evaluation method			9.30% (4 idea units)
Total LF idea units	100% (121 idea units)	100% (84 idea units)	100% (43 idea units)
Total idea units	1227	460	233

The revision was not an easy task for learners, and IADE seemed to reduce the level of difficulty particularly due to its Help Options. This claim is based on 5.78% (7 idea units) coded as “facilitating observed difficulty” and 80.17% (97 idea units) coded as “facilitating expressed difficulty”, both of which marked cases when the learners were struggling to find a solution for a problem they noticed. The former category was used when learners’ utterances made it clear that they were experiencing a difficulty. The latter was used when learners’

actions indicated that they were trying to solve a certain problem. In both cases, access to IADE's Help Options seemed to facilitate learners' problem-solving.

- Facilitating expressed difficulty

  [highlights the first sentence] I cannot think how to revise it. Maybe if I see examples…(Student 28, think-aloud protocol and Camtasia transcript)

  Some sentences I cannot find what's the matter because no matter what I write here, maybe the mind is the same. Let's see the corpus. (Student 40, think-aloud protocol and Camtasia transcript)

  Reads, sighs, mumbles. Says he just can't figure out why this is m1, so he clicks on HO. (Student 54, observation transcript)

  After thinking for a while, asks himself how to revise this part. Says he'll look up the definitions because it helped last time. (Student 65, observation transcript)

  I think it's helps, especially when I don't know what to do. I have no idea. Hmm… definitions helps a lot. (Student 32, interview transcript)

  It's not too hard for me because I had many examples and feedback and definition, so it's ok. It's harder if I do without no help. (Student 61, interview transcript)

- Facilitating observed difficulty

  I think it should be [checks the definitions again] ... (Student 30, think-aloud protocol and Camtasia transcript)

  [thinks, and then opens the definitions again] (Student 44, think-aloud protocol and Camtasia transcript)

  I'm gonna view, I think I'm gonna view the corpus and see what the ... in my discipline the structure of the introduction. [opens an annotated text, scroll down, then goes up to her own text displayed in color.] (Student 53, observation transcript)

  Saying that he doesn't understand the difference between m1 Centrality and m3 Value, he checks the definition of m3 Value. (Student 58, observation transcript)

Next, while the participants paid a lot of attention to IADE's feedback, it did not entirely condition their decision-making regarding output modifications, thus allowing them

the necessary degree of learner control. In some cases, upon resubmission, the feedback returned numbers that did not approximate the average in the field very closely. Students, then, re-read the moves and took time to think about their content, and, if they were satisfied with the structural quality of their discourse, chose not make any changes implied in the feedback. Or, in other cases, having noticed a mismatch between the color-coded feedback and the meaning they put in certain sentences and having carefully reflected about it again, the students decided that it was them who were right, and not the program.

- Learner control

  So, supposedly I choose uh... add more content of m2 because I have only 11.5%, and the average in my field is 21%, but... hmmm... I guess that because of this paper is more uh purpose and not a statement of problem, or specific research… yes. so, it has more m1 and m3, talking about the paper and the background of other research and not much about the problem, so I think it's ok like this. (Student 63, think-aloud protocol and Camtasia transcript)

  The problem is that there are just one main objective in this study. uh... I don't have maybe enough information to add another objective or description in this case. Because it's just... maybe I put a lot of m2 to to present and support the idea. But this is a simple experiment. Maybe... yes, it's not a complex experiment that have many objectives to develop. It's just one objective. And maybe one two objectives to to attain based on the methodology. One or two different results. Not many results there. Yes, I think concie is good enough. (Student 44, think-aloud protocol and Camtasia transcript)

  I will still keep previous research in m1, but try to add something in m3. (Student 62, think-aloud protocol and Camtasia transcript)

  It sounds like it is more a proposal type of paper, so he decides not to worry too much about move because he has no results yet. (Student 28, observation transcript)

  He thinks that "hide" is indicative of m2, but he doesn't want to change because he thinks
  that he is right. (Student 64, observation transcript)

  Explains that by pointing to how problems like this have been solved in his field he is providing background for his study; no way it can be m2. Decides not to change. (Student 44, observation transcript)

> *Why I didn't change color? Because I know better. The software is smart, but it's not high artificial intelligence and it can't know better than me. (Student 44, interview transcript)*

This observation can be linked to 4 idea units (9.30%) in the interview data where the participants were positively appraising IADE's evaluation method. They explained that it was suitable because it was facilitating rather than assessing, which may be one of the factors that allowed them to make their own final decisions. For example, Student 30 said, "I think it good that the software direct me what to pay attention and not say like you're wrong or something. It show color, and I see if I'm wrong. If color is wrong, I'm ok too because I still think again anyway."

In addition, the interviews, like the open-ended survey questions, contained evidence of appropriateness of IADE's feedback for the completion of the revision task. The participants commented on such characteristics as individual and discipline-specificity and iterativity (23.26% or 10 idea units). As seen in 14 idea units (32.56%), the feedback was perceived as suitable for individual students' characteristics such as personality, learning style, or subject knowledge. Some participants implied in one way or another that they needed time to think, read aloud, verbalize their thoughts, consults explanatory materials, and since working with IADE does not impose any restrictions, especially if used out of class, learners could adapt revision with IADE to their usual work strategies. Consider the following examples:

> *The greatest is that is for me personally and about papers of my field. (Student 58, interview transcript)*

> *You know what is nice? That you get right away. Doesn't matter if right or wrong. You know right away and when is like that, of course you want go on and go on. (Student 43, interview transcript)*

For me is good because I can change as much as I want and check myself and don't go ask nobody. I don't like to ask. (Student 28, interview transcript)

I always try draw when I learn something. I remember better if I see. So I really like color in feedback even I don't have to remember. (Student 53, interview transcript)

With definition and examples there I don't learn by heart. I hate that. I like to understand not learn like kid. (Student 65, interview transcript)
I like that I take my time, look here, look there, check, go back. (Student 32, interview transcript)

It's good to work home or any place, any computer. It's like online dictionary, but for me. [laughs]. (Student 44, interview transcript)

6.3.3. Overall evaluation of Learner Fit

To sum up, IADE and its feedback was appropriate for the participants in this study. The summative evaluation of all Likert-scale responses depicted in Figure 6.10a revealed that 46% and 43% of the participants regarded Learner Fit as being an "excellent" and "good" quality, respectively; only 11% thought it was weak. Most of the participants liked using IADE and thought its feedback was comprehensible, appropriate, helpful, and motivating, as summarized in Figure 6.10b.

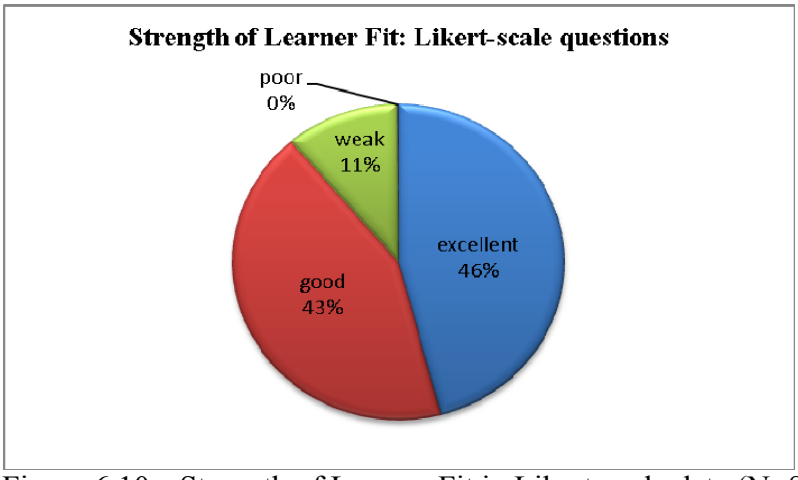


Figure 6.10a. Strength of Learner Fit in Likert-scale data (N=88)

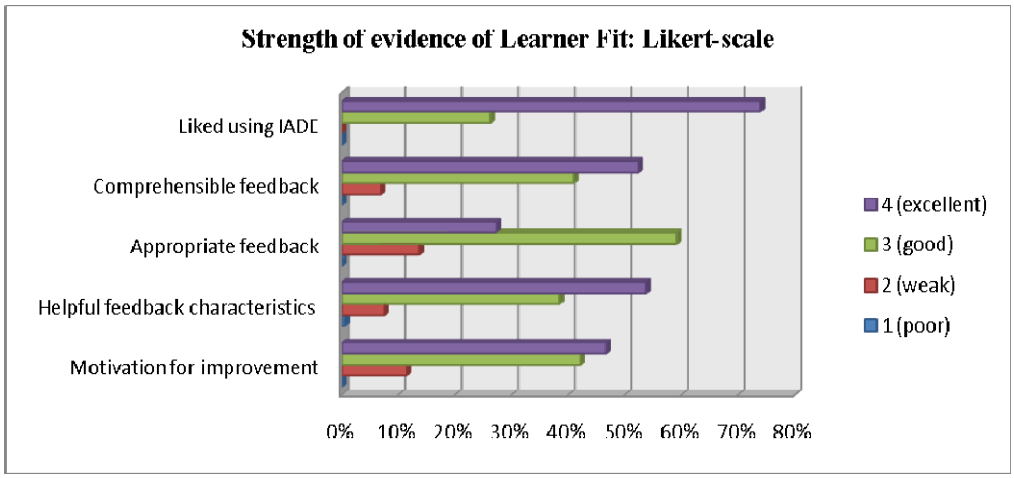


Figure 6.10b. Strength of Learner Fit evidence in Likert-scale data (N=88)

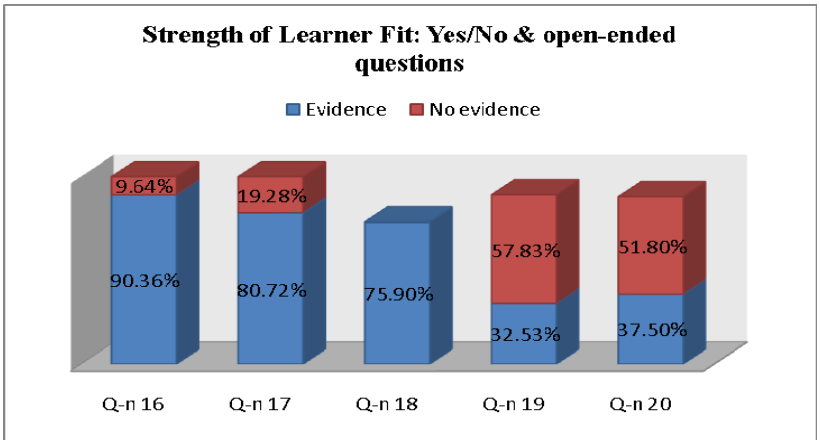


Figure 6.10c. Strength of Learner Fit in yes/No and open-ended survey responses (N=83)

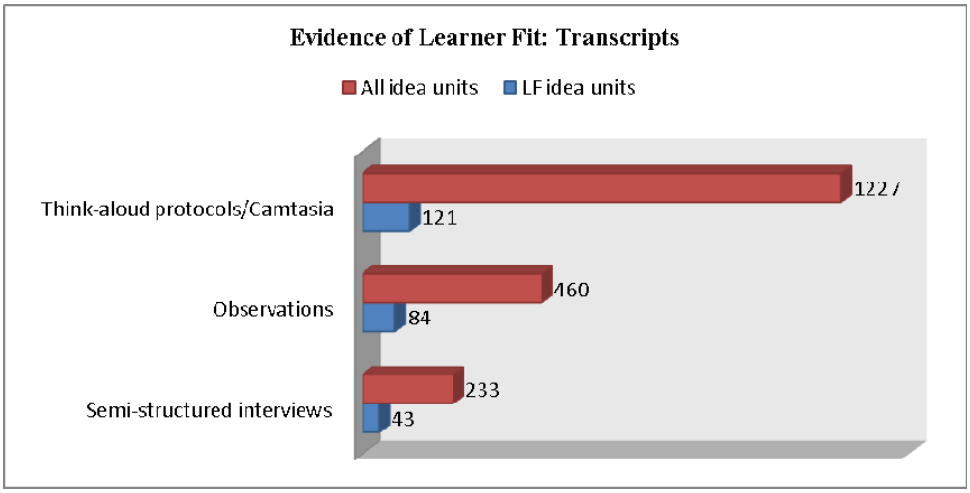


Figure 6.10d. Evidence of Learner Fit in transcripts data (N=16)

The transcripts of introspective and observed data all yielded positive evidence of Learner Fit (see Figure 6.10d above), which not only supported the other findings, but also contributed with additional insights. It appeared that IADE with its Help Options facilitated learners' problem-solving, a claim that may be connected to the findings about their frequent modified interaction. Perhaps, this facilitative feature of IADE enhanced both its Learner Fit and the Language Learning Potential qualities. In addition, considering that one of the characteristics of the targeted students is the fact that they have very deep subject knowledge and much confidence in the arguments they make in their writing, discovering that IADE is also appropriate for them because it allows the necessary degree of learner autonomy without strictly assessing their work products, which are still under development, is another positive evidence of Learner Fit.

6.4. Impact

Impact, i.e. the effects of interaction with IADE and its feedback on learners, was the last CALL quality examined in this study. The same types of data as for Meaning Focus and Learner Fit were analyzed. Evidence of positive impact prevailed in participants' responses to 4 Likert-scale questions and to 5 yes/no and open-ended survey questions as well as in 16 think-aloud protocol/Camtasia recordings, 16 observations, and 16 semi-structured interviews. Evidence of negative effects were found as well. Table 6.11 gives an overview of the overall evidence of Impact, which are related to positive experience, motivation, affect, cognition, and pragmatics.

Table 6.11. Overall evidence of Impact

Data source	*N of part-s*	*Data*	*Positive evidence for I*[6]	*Negative evidence for I*	*No evidence for I*
Survey data: Likert-scale q-ns	88	Responses to:			
		Q-n 14: Did the program motivate you to improve your introduction draft? [motivation]	100%	0%	0%
		Q-n 17: Did you get frustrated when using the program? [affect]	35.23%	64.77%	0%
		Q-n 18: Did revising your draft with IADE have a positive impact on your skills of writing academic introductions? [positive experience]	100%	0%	0%
		Q-n 19: Would you want to use a similar program to practice writing other research article sections? [positive experience]	100%	0%	0%
Survey data: Yes/No & open-ended survey q-ns	83	Q-n 24: Did IADE motivate you to improve your introduction? How? [motivation]	89.16%	0%	10.84%
		Q-n 25: Did you think a lot about how to improve? Why? [cognition]	56.63%	0%	43.37%
		Q-n 26: Were you excited to see improvement reflected in the feedback? Why? [affect]	91.57%	0%	8.43%
		Q-n 27: Would you like to use this program again? When? Why? [positive experience]	90.36%	0%	9.64%
		Q-n 28: Did the feedback influence your revision process or strategies in any way? How? [pragmatics]	90.36%	0%	9.64%

[6] Percentages for Yes/No & open-ended survey question 25 exclude irrelevant responses (15.66%)

Table 6.11. Overall evidence of Impact (continued)

Data source	*N of part-s*	*Data*	*Positive evidence for I*	*Negative evidence for I*	*No evidence for I*
Introspective data:		Transcripts of:			
Think-aloud protocols/Camtasia recordings	16	16 Think-aloud/Camtasia recordings	367/1227 idea units	3/1227 idea units	
Semi-structured interviews	16	16 interviews	40/233 idea units	1/233 idea units	
Observational data:					
Observations	16	16 observations	174/460 idea units	7/460 idea units	

6.4.1. Learner perceptions on Impact

The survey provided quantitative evidence of more beneficial than negative effects. According to participants' choices in Likert-scale question 18, all the students thought that revising their drafts with IADE had a certain degree of positive impact on their skills of writing research article introductions: 39.78% perceived the impact as "very" positive, 39.78% as "somewhat" positive, and 5.68% as "a little" positive (see Figure 6.11).

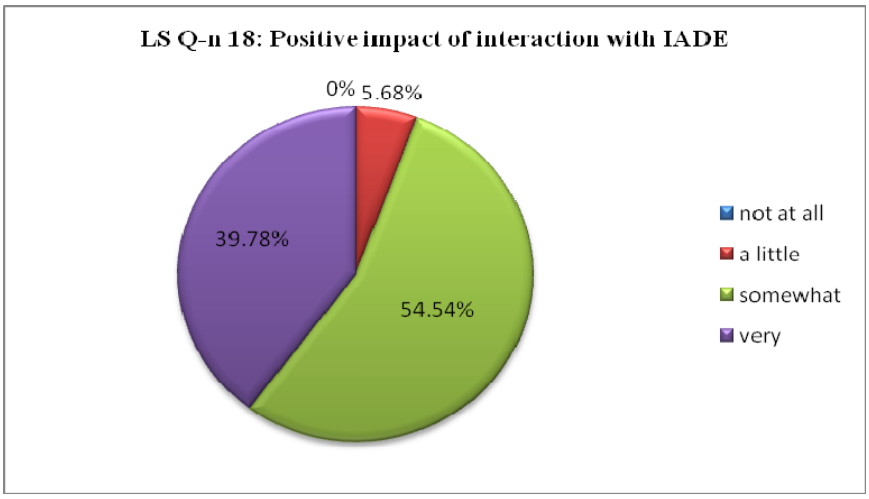


Figure 6.11. Results from Likert-scale question 18

Also, 90.36% (75 students) indicated in yes/no question 27 that they had positive experience with IADE and therefore would like to use this program for future revisions of research articles; some would use it even for conference abstracts, research proposals, and thesis. There were a number of reasons why they would like to work with IADE again: because it was helpful (51.8% or 43 students), because it provided appropriate feedback (25.3% or 21 students), because it enhanced their motivation (4.82% or 4 students), and because of other positive appraisal, e.g., its usefulness, effectiveness, or appeal, (8.43% or 7 students). The few participants who did not see value in using IADE again (9.64% or 8 students), although they did not say that they had a negative experience with it, explained that it was because their particular discipline was not represented in IADE (7.23% or 6 students) or because they did not expect to be writing research articles in the future (2.41% or 2 students). Consider the examples following Table 6.12.

Table 6.12. Results from open-ended survey question 27

	Participants
Willing to use IADE again	90.36%
Reasons:	
Helpfulness	51.8%
Appropriate feedback	25.3%
Enhanced motivation	4.82%
Other	8.43%
Unwilling to use IADE again	9.64%
Reasons:	
Unrepresented discipline in IADE	7.23%
No need to write research articles	2.41%

- Willing to use IADE again

  When I need to write my research articles again, especially those would be published. This program helps me to make my draft seem to be more professional. (Student 27, survey question 27) → Helpfulness

Sure, when I plan to write a paper I would like to try to use it. Because this is the only program I know to help you to write RA introduction. (Student 10, survey question 27) → Helpfulness

When I write other introductions of my papers, I would like to use it to analyze my introduction. At least, I want to know the moves distributions and get feedback for me. (Student 86, survey question 27) → Appropriate feedback

because it actually tells me information very fast and and about my paper and my field. (Student 34, survey question 27) → Appropriate feedback

I would like to use it for each introduction I need to write, and maybe for other parts of the paper. Why? Because I have seen high improvements in my work. (Student 8, survey question 27) → Enhanced motivation

It tell me about my introduction and I know where to fix, so I want to fix especially when I see color change. (Student 20, survey question 27) → Enhanced motivation

When I need to write an article again, I will use this program again because of its usefulness. (Student 57, survey question 27) → Usefulness

Because it is an interesting program. (Student 46, survey question 27) → Appeal

- Unwilling to use IADE again

Not if it doesn't have my area in it. (Student 68, survey question 27) → Unrepresented discipline in IADE

Why I use it with some other field? I don't see why. (Student 87, survey question 27) → Unrepresented discipline in IADE

Maybe not. Because I don't need to write a research paper. (Student 55, survey question 27) → No need to write research articles

Only if I have to, but I don't think so. I want to work in industry and I don't think I write research there. (Student 52, survey question 27) → No need to write research articles

When asked about using a program similar to IADE but developed for other research article sections, none of the participants expressed no desire to work with such a program. As shown in Figure 6.12, the majority of participants would like to have such an opportunity "a

lot" (68.18% or 60 students), 26.14% (23 students) would "somewhat" like that, and 5.68% (5 students) would want that "a little". This is secondary evidence that, overall, IADE and its feedback had a positive impact on learners.

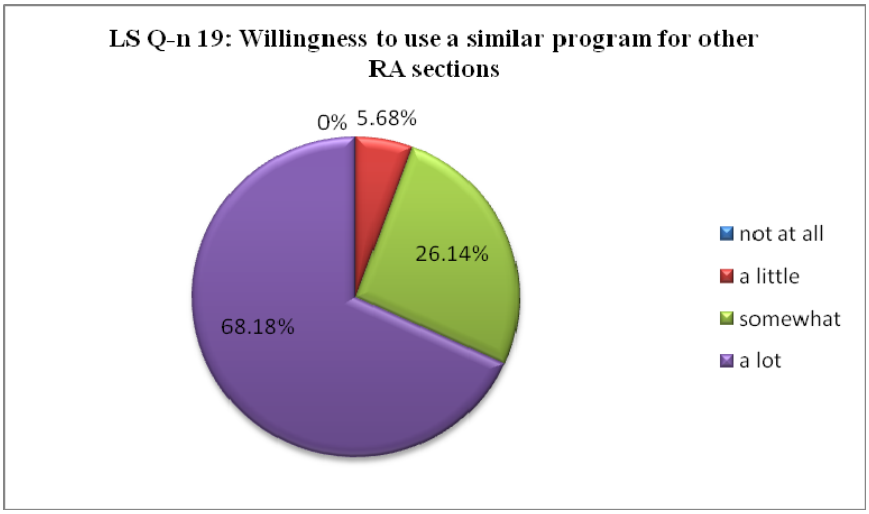


Figure 6.12. Results from Likert-scale question 19

In the questions discussed above, a few students mentioned that interaction with IADE was motivating for them. This opinion gained more solid ground in their responses to scale question 14 and to survey question 24. All the participants in the former source and 89.16% (74 students) in the latter thought that they felt motivated to improve during their revision with the program. More specifically, as graphed in Figure 6.13, it was very motivational for 46.59% (41 students), "somewhat" motivational for 42.05% (37 students), and "a little" for 11.36% (10 students).

The short "no", "not really", "not too much", and answers to survey question 24 provided by 10.84% (9 students) were indicative of neither positive nor negative Impact at the level of intrinsic motivation, and they were not accompanied by explanations either. The "yes" answers were more informative. Table 6.13 summarizes the factors due to which

interaction with IADE was perceived as being motivational by learners. It appeared that the feedback was a major motivational factor as it was mentioned by 40.54% (30 students), who felt stimulated to improve their drafts when they received positive, negative, or guiding feedback, i.e. “feedback constantly direct[ing them] to[wards] improvement” (Student 67). As Student 4 put it, “the feedback really gave me some positive power for the every revision.” The disciplinary orientation of IADE was motivating for 28.38% (21 students) and the opportunity for iterative resubmission - for another 17.57% (13 students). The remaining 13.51% (10 students), did not elaborate on how IADE motivated them, having only stated that it did.

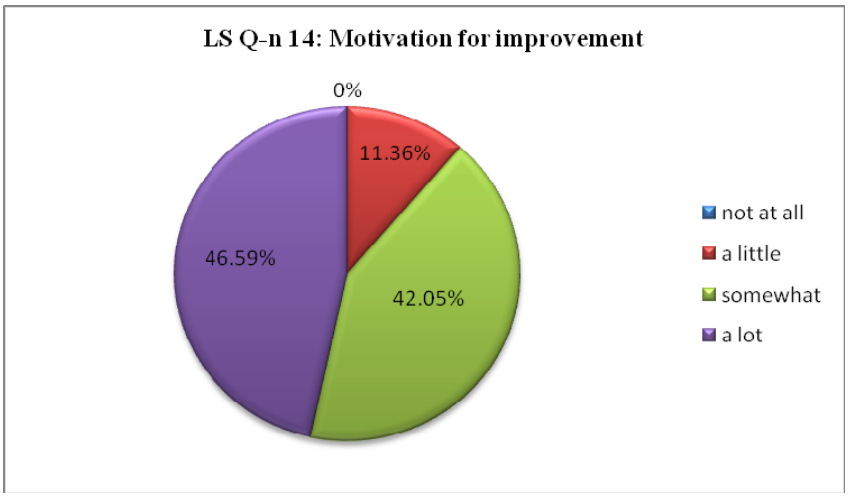


Figure 6.13. Results from Likert-scale question 14

Table 6.13. Results from open-ended survey question 24

Motivating factors	*Participants*
Feedback	40.54%
Positive	26.67%
Negative	63.33%
Guiding	10%
Discipline specificity	28.38%
Resubmission	17.57%
No comment	13.51%

- Feedback

  Everytime I saw that I got better. (Student 83, survey question 24) → Positive feedback

  It motivates me by labeling the moves when change to the right color. (Student 51, survey question 24) → Positive feedback

  It motivated me to revise the sentences and do something to make the proportion of each move better. (Student 52, survey question 24) → Negative feedback

  It let me know that which part is required to fix. (Student 48, survey question 24) → Negative feedback

  I want make the move become what I original thought to get average level and feedback took me step by step. (Student 44, survey question 24) → Guiding feedback

  it gave me some guide for writing the introduction. (Student 4, survey question 24) → Guiding feedback

- Discipline specificity

  because it compared with other corpus in our discipline. (Student 33, survey question 24)

  because it's based on my discipline examples. (Student 8, survey question 24)

  it give me a chance to keep consistent with the norms of my field. (Student 54, survey question 24)

- Resubmission

  IADE *motivated me to improve my introduction through resubmitting. (Student 29, survey question 24)*

  it helped me to revise easy and I was motivate to resubmit all the time. (Student 30, survey question 24)

  Of course because I can submit a lot and see results in seconds. (Student 20, survey question 24)

The survey data also suggests affective impact. Most of the respondents to survey question 26 (91.57% or 76 students) noted that they were excited to see improvement when

the feedback on their modifications was returned. For instance, Student 38 wrote, "When I saw my improve feedback, I feel good" and Student 65 remarked that he "did get excited, not like a child ;-) but as a happy student." Although 36.84% (28 students) did not explain the causes of their excitement, those who did thought that the factors listed in Table 6.14 generated this emotion. For 30.26% (23 students) it was the feeling of accomplishment, which was oftentimes not easy to achieve. In other words, having gone through a series of trial and error attempts to improve their drafts, the learners felt like their hard work paid off when they saw a positive change in IADE's feedback. For 23.68% (18 students) it was the proof of improvement displayed in IADE's feedback that had this effect. A few students (9.21% or 7 students) experienced disappointment due to multiple unsuccessful output modifications, and they felt like their excitement was stronger after that.

Table 6.14. Results from open-ended survey question 26

Causes of excitement	*Participants*
Feeling of accomplishment	30.26%
Proof of improvement	23.68%
Previous disappointment	9.21%
No comment	36.84%

Here are a few examples from the responses to this question:

- Feeling of accomplishment

  I had the feeling that I was achieving my target (Student 75, survey question 26)

  It seems that I achieved something. (Student 41, survey question 26)

  I do a lot to improve it, so I excited of course. It means "I made it"! (Student 93, survey question 26)

- Proof of improvement

  because it means my English interpretation ability has improved. (Student 53, survey question 26)

I was excited when I was in the average. I felt happy because I know that I was in a good point of the introduction. (Student 47, survey question 26)

Why? It is just good to see improvement. (Student 27, survey question 26)

- Previous disappointment

I was excited, especially after I was sad. (Student 29, survey question 26)

Not always because it didn't happen very often. Sometimes I try too many times and nothing. But I was really excited when I finally got it right. (Student 68, survey question 26)

However not always. It upset me if it was the same when I submit again and again. But sometimes after I'm upset I get so happy! (Student 73, survey question 26)

As the last examples above suggest, some students' experience with IADE was not entirely positive. Responding to Likert-scale question 17, 3.41% (3 students) indicated that they were very frustrated during the interaction with the program, and 25% (22 students) were "somewhat" frustrated. However, the rest of the participants described their experience as either "a little" frustrating (36.36% or 32 students) or "not at all" (35.23% or 31 students). Overall, 64.77% of respondents pointed to negative affective Impact.

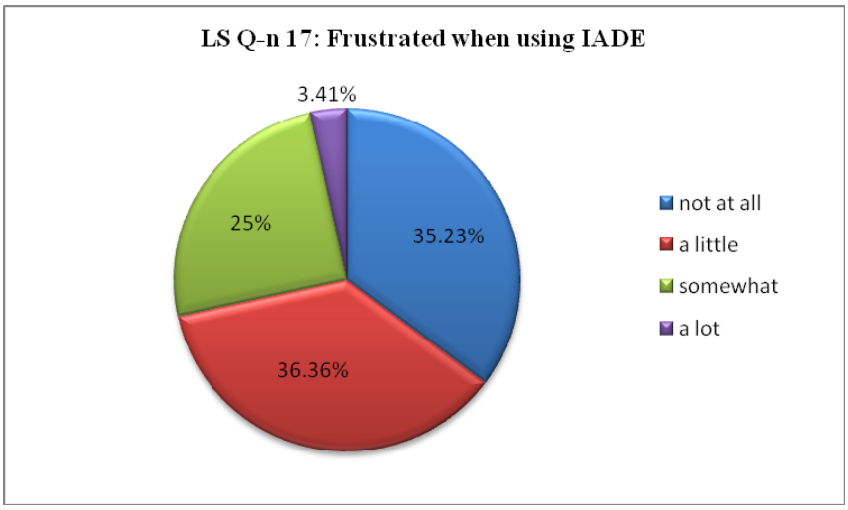


Figure 6.14. Results from Likert-scale question 17

Another effect, which was at the level of pragmatics, became evident in participants' answers to survey question 28, where 90.36% (75 students) acknowledged that the feedback influenced their usual revision process and strategies. The other 9.64% either were not sure (4 students) or did not think so (4 students). The rest of the students provided comments that revealed how their revision strategies were indeed altered. Some of the participants (6.67% or 5 students) pointed out that they did not have any revision strategies to start with and that it was difficult for them to express an opinion from this perspective. Others, who were aware of how they were used to revising their writing, referred to acquiring new strategies. One of those strategies was learners' reliance on the guidance of IADE's feedback (18.67% or 14 students). In other words, they took the feedback as the only guidance – if it said that a move was in the average range compared to their discipline, they would not work on it anymore, and if there was some negative evidence about a move (especially numerical), they would try to improve on that particular move.

A similar strategy, which was articulated by 10.67% (8 students), was to set a goal – the goal of reaching the standards of the discipline as presented by the average percentages in the feedback. While setting goals is a positive choice, it is still reliance on the numerical feedback that leads to accomplishing this type of goal with IADE, which is why this strategy is also somewhat confining. Additionally, learners' focus on discourse form also became part of revision for some participants (9.33% or 7 students) who mentioned that they consistently paid attention to the distribution of the color-coded moves. A larger group of participants (26.67% or 20 students) developed their focus on meaning into a strategy, by which they intentionally thought of the functional meaning of their sentences. For example, Student 89 explained, "I decided to check every sentence and think what step it function, so I checked if

the color was right, then good, if not then I know I have to think about what I say and how make it sound like what should be." It must be mentioned here that for 8% (6 students) the focus on meaning strategy took the form of "pay[ing] much attention to wording" (Student 31) and of move-specific vocabulary search and use. (See Table 6.15)

Table 6.15. Results from open-ended survey question 28

Revision strategies	*Participants*
Reliance on the guidance of feedback	18.67%
Setting the goal to reach a standard	10.67%
Focus on discourse form	9.33%
Focus on functional meaning	26.67%
Move-specific lexical output modifications	8%
No previous strategy	6.67%
No comment	28%

- Reliance on the guidance of feedback

  I revised base on the feedback advices. (Student 91, survey question 28)

  the feedback influenced my revision strategies in most way. I revised like the feedback tell me. (Student 70, survey question 28)

  The feedback give me directions how to revise my drafts. (Student 57, survey question 28)

- Setting the goal to reach a standard

  it made me meet the standard portion of moves. (Student 7, survey question 28)

  I notice that I would want the desired percentages. (Student 75, survey question 28)

  I added more for moves to reach the percent. (Student 30, survey question 28)

- Focus on functional meaning

  It was very different than usual. I usually just read one time and if I see mistakes I fix them. Now I consider color first and begin to think of what step my moves are. (Student 82, survey question 28)

I want make the move become what I original thought (Student 44, survey question 28)

It made me pay attention to how I say my ideas because they not always sound like they should. They sometimes sound like for another purpose (Student 83, survey question 28)

- Focus on discourse form

I always tried to get a balance of the proportion of each move. (Student 85, survey question 28)

I started to think how clearly indicate how I organized their introduction. (Student 72, survey question 28)

I think about the balance among the moves and the structure. (Student 27, survey question 28)

it made me think what words I should need to use correctly in different move. (Student 4, survey question 28)

- No previous strategy

I didn't have strategies. I don't know, maybe I have a strategy now. (Student 55, survey question 28)

I don't have a revision process before I use the program (Student 59, survey question 28)

I don't know. I never think I have to have strategy. (Student 36, survey question 28)

The finding that the participants made focus on the meaning and on the form of the discourse part of the revision process is promising, especially considering that these types of focus contribute to learning and improvement. However, some learners' heavy reliance on the program's automated feedback, their desire to reach average percentages in their field, and too much attention to move-specific phraseology are much less encouraging findings. These strategies are limiting in that they prevent the development of learner autonomy, which would allow learners to make decisions for writing that will not be facilitated by IADE or

any other CALL application. Moreover, what is also lamentable is that learners may not self-evaluate their work products; rather, they may be more likely to take the automated feedback as the only indication of the quality of their writing, trying to only address the salient negative evidence in the feedback and end their revision as soon as the feedback would display percentages close to the average in their discipline.

Perhaps, these may have been the reasons why 24.1% (20 students) indicated in their responses to survey question 25 that they did not think about how to improve their draft while revising it. Student 50 wrote, "Not a lot, only a little. I know what I lack and then to add the lack" and, similarly, Student 33 remarked, "No, just tried to meet the feedback." Unfortunately, not all the participants addressed this question; 15.66% (13 students) provided brief irrelevant responses and 3.61% (3 students) were not sure how to answer, which prevents from making confident claims. Nevertheless, the responses of 56.63% (47 students) who said that they did think about how to improve their writing shed some useful light on this issue. Interestingly, their explanations were directly related to the newly acquired revision strategies that included conscious focus on discourse meaning and form. Consider the following quotes:

> *I thought a lot to express the meaning more precisely. (Student 52, survey question 25)*
>
> *because I need to compare my thought with what is in* IADE *(Student 30, survey question 25)*
>
> *Because sometimes it took me quite a long time to find appropriate words to express my idea. (Student 25, survey question 25)*
>
> *I tried to identify the patterns of moves and steps (Student 61, survey question 25)*
>
> *If I don't think a lot about how to improve, I can't get real feedback about my structure from every revision. (Student 4, survey question 25)*

It seems that those learners who focused on the form and on the meaning of their draft found themselves thinking more often than they normally would. Some of them believed that the more they revised, the more deeply cognitively engaged they became. In Student's 72 words, "Actually not at the beginning, not when I began to revise, but a lot later after I figured my idea didn't match with colors." Others explained that it was necessary to think a lot on their own "because there is no best, only better" (Student 27), possibly meaning that having reached average percentages did not necessarily indicate that the draft was perfect.

6.4.2. Introspective data on Impact

Introspective insights on Impact were gained from 16 think-aloud/Camtasia protocols, observations, and semi-structured interviews. In think-aloud/Camtasia data, 143 idea units were coded for Impact, with an average of 8.94 idea units per participant. In observations, the number of Impact idea units amounted to 181, with an average of 11.31 idea units per participant. The interviews yielded 43 idea units with evidence of Impact, with an average of 2.69 idea units per participant. These data sources substantiated the survey findings with evidence that the participants experienced influences at different levels: affective, intrinsic, pragmatic, and cognitive, which are summarized in Table 6.16.

Table 6.16. Themes indicative of Impact identified in transcripts

Impact levels and themes that emerged	*TA/Camtasia*	*Observations*	*Interviews*
Cognitive			
Cognitive involvement	38.65% (143 idea units)	25.41% (46 idea units)	30.23% (13 idea units)
Pragmatic			
Strategy development	34.32% (127 idea units)	31.49% (57 idea units)	11.63% (5 idea units)
Influence on revision process			9.3% (4 idea units)

Table 6.16. Themes indicative of Impact identified in transcripts (continued)

Impact levels and themes that emerged	*TA/Camtasia*	*Observations*	*Interviews*
Affective			
Emotional involvement	7.84% (29 idea units)	6.63% (12 idea units)	
Positive learning experience	7.84% (29 idea units)	12.71% (23 idea units)	37.21% (16 idea units)
Negative learning experience	0.81% (3 idea units)	3.87% (7 idea units)	2.33% (1 idea unit)
Intrinsic			
Motivation	10.54% (39 idea units)	19.89% (36 idea units)	9.3% (4 idea units)
Total I idea units	100% (370 idea units)	100% (181 idea units)	100% (43 idea units)
Total idea units	1227	460	233

The think-aloud/Camtasia recordings and the observation notes captured many instances (38.65% and 25.41%, respectively) when the learners appeared to be cognitively involved. In 30.23% of interview transcripts idea units, the participants stated that they spend much more time on revision because they had to think more than usual. They engaged in self-reflection, self-analyzed their output, explained their intent, confirmed or rejected own hypotheses, and drew conclusions based on IADE's feedback and materials in Help Options. In addition to being overtly expressed in students' utterances, this cognitive involvement was also noticeable when they engaged in deep thinking reflected in long pauses, interjections like "hm", and body language as well as when they highlighted and re-read parts of their writing. What is interesting, although not surprising, is that thinking was more called for when the learners noticed negative evidence related to intended functional meaning and the form of their discourse. Following are a few examples from the data:

[highlights the next sentence and reads carefully] I'm wondering why the previous research papers do not have that. Maybe they just don't include hypothesis in my field? I guess so. (Student 27, think-aloud protocol and Camtasia transcript)

[thinks. whispers quietly] I have no idea why it's... they are classified as m1. [opens definitions for m1] it's not the centrality... topic generalization... they are also not the general statement. Obviously, they are not belongs to previous research. [back to the color-coded feedback. reads the sentence again, trying to figure out what it's function would be] (Student 29, think-aloud protocol and Camtasia transcript)

[opens the paper materials again] this part is m3. I think it's about this paper. This is ... may include I think... ummm, and this is... [thinks] ummm, I think this part of this paper is generalize the model is developed. This is... [thinks] (Student 54, think-aloud protocol and Camtasia transcript)

Thinks about the content and coherence. (Student 43, observation transcript)

Decides that she has to add some more m3 Description of current research, but she says that she has to think about it more. (Student 43, observation transcript)

Goes on reading very attentively. (Student 64, observation transcript)

I know! It take me long to think. Maybe because it's hard to figure out how to say the gap or hypothesizing and so on. (Student 58, interview transcript)

Yes. I think a lot now. (Student 44, interview transcript)

I read what I write again many times because I think what's wrong. And then I think how to correct. But it's hard to know what's wrong. (Student 30, interview transcript)

In terms of pragmatic influence, the transcripts attested to learners' development of new revision strategies during the interaction with IADE. The Impact idea units coded as strategy development were second in frequency after the cognitive involvement category. The emerging strategies were very similar to the ones mentioned by the participants in their open-ended survey responses - following the feedback and focusing on meaning and form. To a certain extent, they were determined by the program's features, which offered opportunities for resubmission, immediate feedback of numerical and color-coded types, and

availability of Help Options. While making use of these features was common for all the observed participants, the sequence and frequency with which they were used changed gradually as the revision process unfolded. Reliance on feedback and resubmission were more frequent at the beginning as opposed to cognitive involvement and consultation of help, which became more prominent later on.

All the students began revising by trying to improve on the move that was the farthest from the average in their discipline. For example, Student 30 said, “OK. Because I have average about 55% of m1, I have 61% roughly, so it's ok for me. And for m2, the average is about 14, and I have 14%. And for m3 the average is 30%, and I have so little. It’s ok. I think I just have to add more move 3” (think-aloud/Camtasia transcript). Later, the students developed a more thoughtful approach and acted more based on their self-reflection, constantly accessing the Help Options and comparing the definitions and examples of moves and steps with their own draft. For some, reaching this stage took more time than for others because they appeared to continue focusing mainly on the numerical feedback, taking actions only when it displayed some negative evidence and attending to that by making quick minor modifications, which were resubmitted multiple times. The excerpts of raw, uncoded transcripts quoted in Figure 6.15 shows how different the nature of Student’s 29 output modifications was initially and subsequently.

Initial revisions:

Too much move 1. Let's see if I change the position, what will happen.

[goes down to the revision box and highlights part of his m1 text.] Let's change some of the previous research review into later. [moves the highlighted part to another place in the text]

Figure 6.15. Initial and subsequent revision strategies (think-aloud/Camtasia transcript, Student 29)

OK, I already changed it, changed the location of it.

Hmm... Although I changed the location, it is still regarded as m1. I'll try to move it to the end of the paragraph. [cuts and pastes that part to the end of the introduction, then submits]

All right, now I know that it is not because of the location. Maybe it's because of the language since I already changed the location of this.

[goes to the Help Options, annotated corpus, and looks at one annotated introduction] Oh, just a quick comment. I think another way to revise it would be to add some parts like the structure or the summary of the outcomes.

Now I've decided to add something. Now I'm focused on the last paragraph of my introduction. I'm trying to add something.

Subsequent revisions:

The last sentence is recognized as m1, but I meant it as m3. So, I'm looking at it and see what happens. [reads the sentence]. It doesn't show that.... hmmm. It should be the principle result of the research, but I don't know why the program detected it as m1.

[reads and thinks. looks up examples in the annotated corpus. Goes back to the color-coded feedback] I'm still looking at the last sentence ... and to see if I can make some changes and the program can respond to these changes. Maybe I can try to be specific and say that "the results of this paper". Yeah, let's try to do that. [makes the change" "The results show that..." --> "The results of this study show that..." and submits]

Yep! All right! Now it changed to m3, it's what I mean. So that's good!

Figure 6.15. Initial and subsequent revision strategies (think-aloud/Camtasia transcript, Student 29) (continued)

Another important observation is that focus on meaning seemed to be a determining factor in the development of new, more effective revision strategies. Figure 6.16 below captures the point where Student 61 transitions from a sporadic to a more thoughtful revision strategy that involves focus on functional meaning, more cognitive involvement, and more effective use of Help Options. One type of available help, the revision tips, particularly facilitated such a transition. Those students who consulted them began to focus on meaning

earlier in the process. However, even those who did not access that kind of help gradually developed strategies that included meaning focus.

I can move it right before the last sentence. [cuts the sentence and places it later in the text.] well... ok, let us see the result again. [re-submits]

nothing happened! [looks at the color-coded text. thinks] maybe I can delete this... this words [highlights "overcome its disadvantages first"] so... this should help [deletes the "overcome its disadvantages". resubmits]

oh... average of m3 is 6.45. [looks back at colors. does to the revision box and adds: "First, the double-notch structure device is developed." re-submits]

[looks at the modified sentence in the color-coded feedback. the color is still blue] uh... structure... [goes down to the revision box. cursor over the sentence. thinks] so, I was confused about the program. Uh, I was trying to make every... every move of average length. But I can't. So, maybe I can add something to m3.

Transition:

So, which is in m3? [looks at the green part of his text. opens HO step stats. then the definitions] summarizing methods... outline the structure... yeah. research questions... summarizing method... announcing principle outcomes... let me see some examples. Yeah.

[opens the AC. reads m3, cursor over green text] we identify... [indistinct, reads in quiet voice, cursor over] oh! opens another annotated text, finds green text, highlights parts of it] present tense... yeah... [highlights "is developed" in his text. thinks. reads in half voice. highlights "was developed". sighs deeply.

[clicks on green sentences in AC to see the steps] let's see more example. [opens another annotated text and clicks on green sentences] definition m3. outlining structure... we have organized this paper. here... this paper... yeah. [checks the definitions in the HO] so, they have a... they have strong words to indicate that the research results belongs to this paper.

[reads the sentence he has previously modified and adds "in this paper" at the end] let me try again. [resubmits]

It’s great! And it is still below average in my discipline. Uh, I need to add something else because in m3 I only outline the structure, but nothing else. So, m3. [looks at the colored text]

Figure 6.16. Transition from sporadic to consistent revision strategies (think-aloud/Camtasia transcript, Student 61)

move 3 has other steps. [opens the HO step stats] announcing the purpose... [opens definitions with the few examples provided there] the aim of this study was bla-bla... the purpose of the present study was...

ok. What I wanna do... [reads his own text meant as m3] the special structure... um... [thinks] what... [mumbles] yeah... [deep sigh] know what... um... [highlights a sentence] this is the first step. And then...

[changes: "First, the special structure and analytical results of DTMCPS was discussed. Then, the double-notch structure device is developed in this paper." resubmits]

[blue sentences changed to green] Yeah, the program becomes smarter. First, then... [reads from the blue sentences left meant as m3] in fact... after that... yeah. That's what I want to say. [looks at the numerical feedback] and... m1, m2, m3. All about average in my discipline. That's great!

Figure 6.16. Transition from sporadic to consistent revision strategies (think-aloud/Camtasia transcript, Student 61) (continued)

Although the participants did not finish revising their drafts when being observed and recorded, towards the end some of them reached a stage where their strategy consisted of a very detailed self-verification. In short, they began checking the colors in the feedback against the intended functional meaning sentence by sentence, thus obtaining a better sense of how their introduction unfolded overall and ensuring the quality of their drafts. Figure 6.17 contains such an example from the transcript of Student 30.

So, right now the structure at move level seems to be fine because all my moves are about average. And how I check the moves for the presence and quality of each step. ok.

[opens HO definitions and reads. then looks up at the colored text]

The first sentence is m2 because m2 aims at showing the reader the existing gaps or problems and I did mean to introduce the previous work, so I think m2 is ok because aims at showing the existing gaps or problems, and my work is based on our previous work. So, I first do the introduction of our previous work, refer to the first paper, which works on this problem, and my work is...

Figure 6.17. Transition from sporadic to consistent revision strategies (think-aloud/Camtasia transcript, Student 30)

I use this solution proposed in the first paper to solve a real problem in our life. So, the second sentence is m1. M1 is establishing a field. In these sentences I introduce this specific field in which our previous solution can be used. So, I think for these sentences m1 is ok because for dynamic MRI this method can be used.

And this is also m2. [checks m2 in definitions] Yeah, in this sentence I mean to say why our method is helpful for such applications because the MRI image is time-consuming and expensive to take completely samples. So we use this method to reduce the samples we need to take. So I mean to demonstrate that this specific topic, in other words, this method is helpful to this. So, I use m2 to uh... highlighting a problem or indicating a gap? [checks definitions] Yeah, this is m2, step 2, highlighting a problem. And then [reads and highlights a sentence in the colored text] yeah, this is some other works which also aim to solve this problem using different methods than us, and this sentence points out this existing method is a batch solution, not sequential, and so that ... that's an advantage of our method. So it's m2, red one.

I think it should be [checks the definitions again] ... also step 2 of m2, highlighting a problem because it illustrates limitations of existing works. And this one, highlighting a... this one, the red one.

And after that, this... "on the other hand..." bla bla... this is... [highlights a sentence] this gives some general information of our method, why it is better than existing methods which I introduced above.

And then after that I give the some general discussion of... I give some specific work in this paper. [mouse over green text] For example, I outline the work I do in this paper. We developed a method, then second we use the result of such things to test our algorithm, and it is ... m3. [thinks, and then opens the definitions again] Yes, I think it belongs to summarizing methods and announcing principle outcomes.

In the last sentence I indicate that the simulation result is given in section 4, in picture 1b. So, it is uhh... outlining the structure... not outlining the structure of the paper. It's also m3 announcing principle outcomes. So, that's the structure and the content of introduction section. [keeps checking the m3 steps' definitions while deciding and confirming the move steps for her sentences]

[goes back to the colored text, highlights part of the green text] So, I revised here.

Figure 6.17. Transition from sporadic to consistent revision strategies (think-aloud/Camtasia transcript, Student 30) (continued)

To summarize, an interesting pattern became visible in the think-aloud/Camtasia and observation transcripts. As shown in Figure 6.18, it appeared that, at the initial revision stage,

when the learners paid attention mostly to the numerical feedback, their modifications, such as change of placement, deletions, and substitutions of words, were sporadic, inconsistent, and not very successful. As a result, they did not seem motivated and experienced frustration. Further, once the learners began focusing on functional meaning, they tended to reflect on and revise sentence by sentence, or, when they focused more on the form of their discourse, they tended to revise move by move. In both these cases, the revision process became more organized and sequenced, resulting in more successful output modifications, which were accompanied by more bursts of positive emotions.

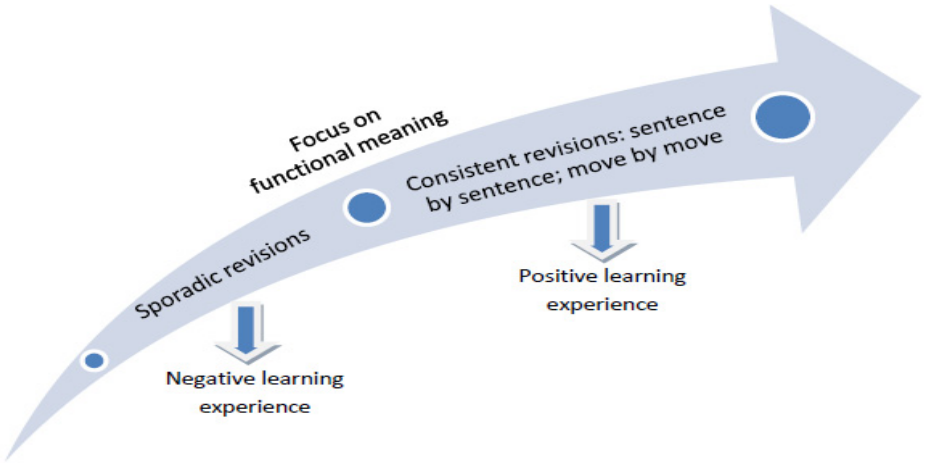


Figure 6.18. Participants' revision process

Affective impact was noted in the recordings of 12 out of 16 observed participants, particularly when excitement was verbalized. In many cases, without getting emotional, the learners simply expressed satisfaction when receiving positive feedback upon resubmission or when self-confirming successful output modifications, which spoke of positive learning experience. On the other hand, their dissatisfaction when seeing negative feedback, which was often returned for repeated attempts to modify the same piece of writing, was indicative of negative learning experience. Positive experience appeared to be more frequent compared

to negative (see Table 6.16), possibly because the learners realized that improvement does not come easy and that "it's normal that I don't get right the first time" (Student 43, semi-structured interview).

> *Oh, I know! In m2 if I want... there is a step... indicating a gap. In this step describes some problem in the previous research. Oh, I know! It's natural! (Student 40, think-aloud protocol and Camtasia transcript)* → Emotional involvement
>
> *Oh! that looks nice! A lot of m2 appears. (Student 43, think-aloud protocol and Camtasia transcript)* → Emotional involvement
>
> *Plays with the sentence a little bit, and the last sentence is successful. Gets very excited. (Student 65, observation transcript)* → Emotional involvement
>
> *Yes, the part that I add in m2 works in the m2 now. ok. (Student 44, think-aloud protocol and Camtasia transcript)* → Positive learning experience
>
> *Yeah, now finally it's m3. Yeah, it's good. (Student 29, think-aloud protocol and Camtasia transcript)* → Positive learning experience
>
> *To see how colors change and average change is so good for me because I know I did good. (Student 32, interview transcript)* → Positive learning experience
>
> *Too much needs to be considered in writing! [he sounds overwhelmed] (Student 65, think-aloud protocol and Camtasia transcript)* → Negative learning experience
>
> *She obviously needs guidance. She struggles, is disappointed, doesn't know what else to do. (Student 30, observation transcript)* → Negative learning experience
>
> *I want to give up after I submit so many times. (Student 32, interview transcript)* → Negative learning experience

Next, it appeared that the learners gradually developed an intrinsic desire to improve and that their motivation was driven by IADE's both negative and positive feedback. Negative feedback motivated them to address their problems while positive feedback motivated them to continue revising. However, if negative feedback was too frequent, meaning that the modifications made were repeatedly unsuccessful, the participants were

more likely to become frustrated and either give up or decide to return to that modification later. What also seemed to have a motivational effect was the content in IADE's Help Options, which the participants accessed when struggling with finding a way to approach certain deficiencies in their drafts. The interviews revealed that the iterative and instant characteristics of the feedback were also motivational because that maintained the students' interest in verifying the effectiveness of their modifications. Student 32 clarified, "It's helps because what I change is still fresh and it make sense when I have feedback right away. When I get some back from the teacher, I forget and don't worry about every comment too much" (interview transcript). It also seemed that the more cognitively engaged the participants were and the more effective revision strategies they employed, the more motivated they became.

> *So, there is more room for improvement. (Student 27, think-aloud protocol and Camtasia transcript)*
>
> *[the modified parts turned into green as intended] that's really what I want. Again! [laughs] everything is about average. ok, the number is good, the length is good. ok. hm. let me see what other things I need to incorporate in m3. (Student 43, think-aloud protocol and Camtasia transcript)*
>
> *Now everything is about average again. Keeps thinking about what to add and how to make m3 stronger. (Student 62, observation transcript)*
>
> *She immediately says that it's not what she means. Decides that she definitely needs to revise. Moves closer to the computer and takes a position for focused work. (Student 28, observation transcript)*
>
> *I will go home and continue because the software gives me feedback and also many examples. (Student 53, interview transcript)*
>
> *You know, I don't think I add more if not this feedback about my field. (Student 61, interview transcript)*

> *Most of the time when I check if I have right move, especially if I think about steps, I see the good color when submitting. So of course it make me happy and I want to go on because I know what to do to get better. (Student 30, interview transcript)*

Given these findings, it may be inferred that, during revision with IADE, the learners experienced impact at different levels, which seemed to have had certain effects on one another (Fig. 6.19). The depth of cognitive involvement conditioned the use of revision strategies, thus influencing the pragmatic aspect of learners' revision process. Specifically, when the learners almost entirely relied on the numerical automated feedback, the degree of cognitive involvement was low, and, therefore they made superficial and ineffective output modifications. Later in the process, when the learners started to focus on functional meaning due to the color-coded feedback, the degree of cognitive involvement increased, and, therefore they made consistent and more successful changes to their drafts. These pragmatic choices, in turn, had either positive or negative effects on learners at affective and motivational levels.

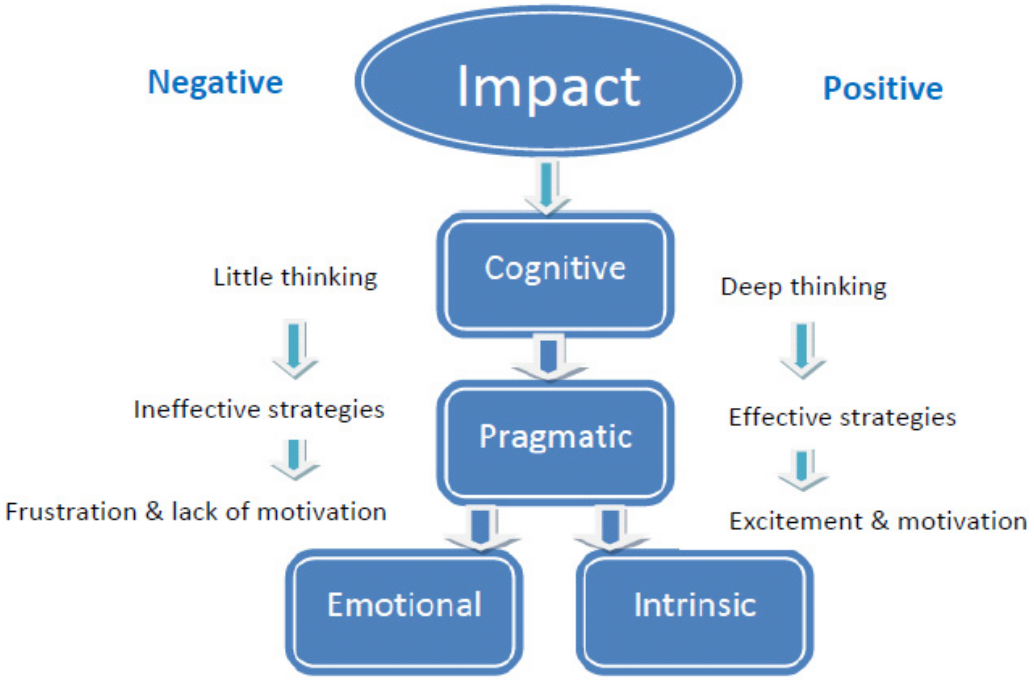


Figure 6.19. Levels of negative and positive Impact

This inference, in turn, suggests that the two types of feedback, numerical and color-coded, although both perceived as beneficial by the participants, may have had opposite effects on learners' approach to revision. The former may have inhibited the effectiveness of the revision strategies they employed, possibly encouraging students to limit revision to approximating percentages. The latter, on the contrary, seemed to enhance this process by facilitating meaning focus. Considering the observed effect of chosen revision strategies on participants' motivation and emotions, it may be presumed that the type of automated feedback is directly related to the type of impact experienced by learners.

6.4.3. Overall evaluation of Impact

Strong evidence of positive Impact were found. Likert-scale responses largely pointed to learners' positive experience with IADE, averaging 52% of "excellent", 37% of "good", and 11% of "weak" evidence (Fig. 6.20a). According to the more detailed survey findings summarized in Figures 6.20b and 6.20c, most of the participants were motivated to improve, experienced positive impact, felt excited about their improvement and sometimes frustrated about lack thereof, and expressed willingness to work with IADE in the future.

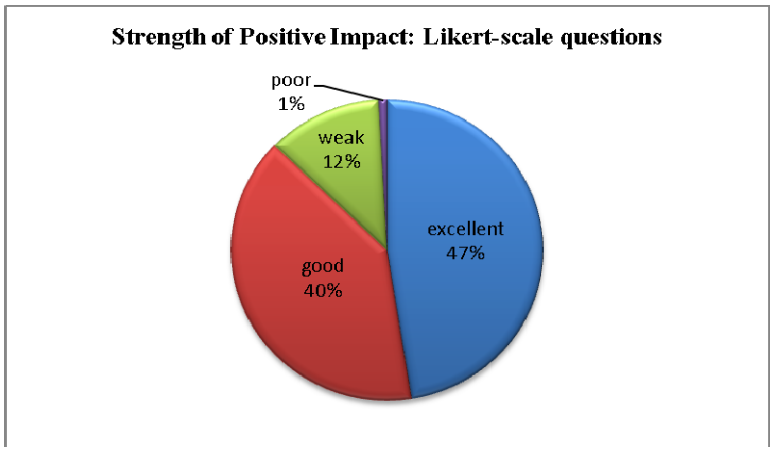


Figure 6.20a. Strength of positive Impact in Likert-scale data (N-88)

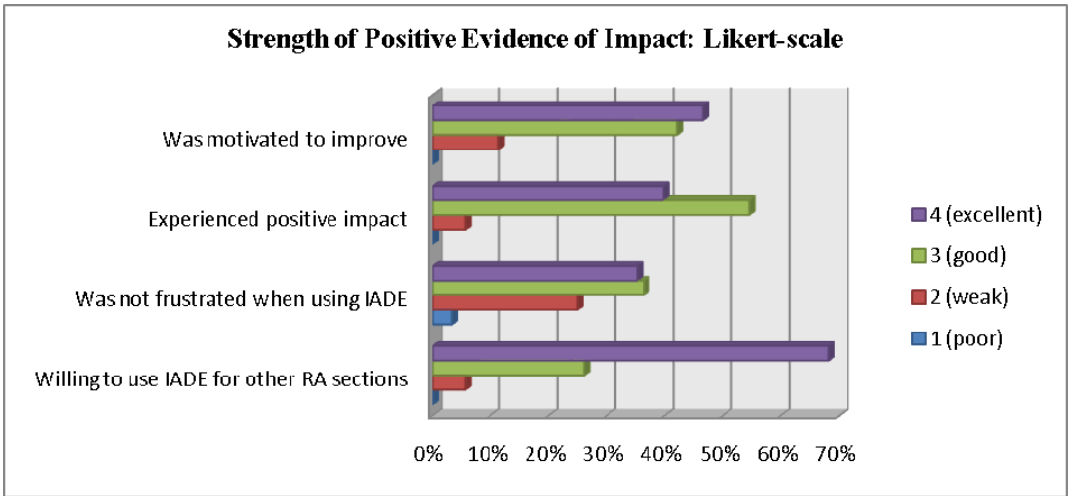


Figure 6.20b. Strength of positive evidence of Impact in Likert-scale data (N-88)

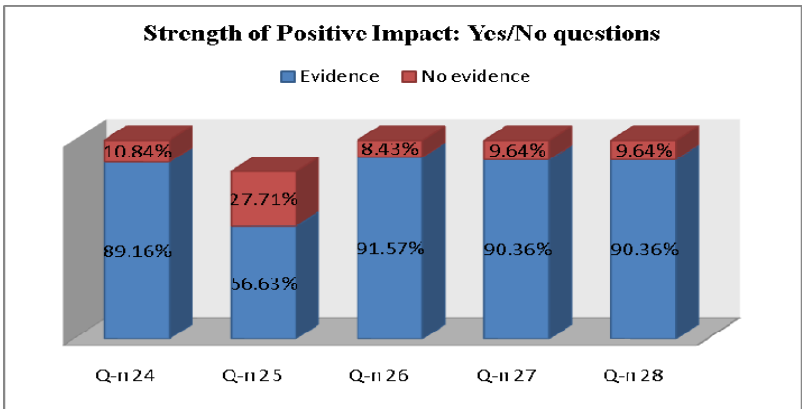


Figure 6.20c. Strength of positive Impact in Yes/No survey data (N=83)

Similar evidence was found in all transcripts data (Fig. 6.20d), which helped better understand how learners' revision process was influenced and what consequences this impact had. In short, the revision process appeared to gradually change during learners' interaction with IADE. The strategies they used during revision evolved from sporadic changes based on the feedback to more consistent output modifications based on focus on functional meaning. This change was determined by the degree of cognitive involvement stimulated by the numerical or color-coded feedback and had an increasingly possitive effect on learners' at emotions and motivation.

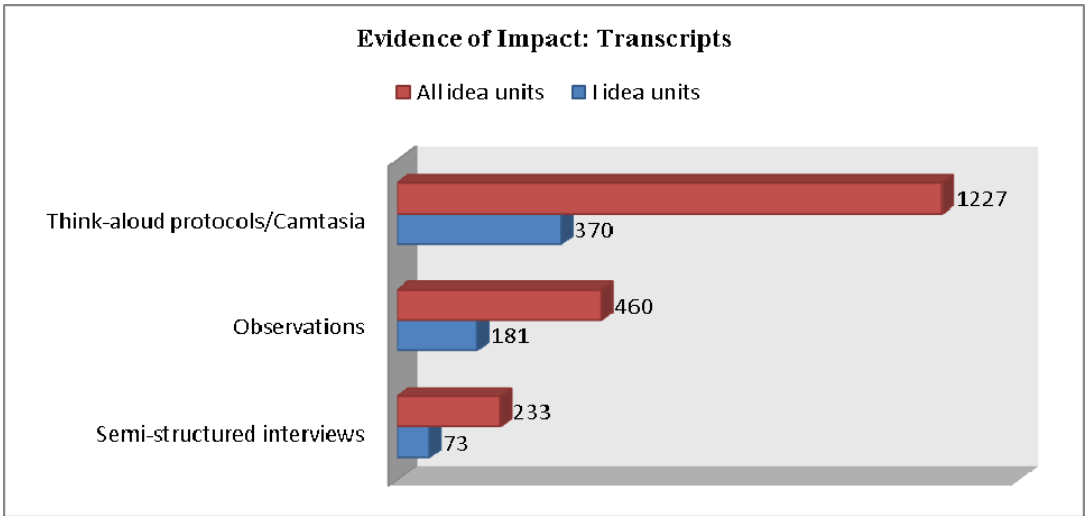


Figure 6.20d. Evidence of positive Impact in transcripts data (N=16)

The findings on positive Impact can be linked to some aspects of other CALL qualities investigated in this research. Most importantly, they confirm the value of Meaning Focus, which, when present, not only contributes to learning but also to a more positive learning experience. Also, the observations that, towards the end of the revision process, the participants tended to rely more on their own thinking and reflections rather than on the numerical feedback may reinforce the claim that IADE possesses Language Learning Potential because it fosters enhanced understanding and adequate Learner Fit because it encourages learner autonomy.

6.5. Chapter summary

This chapter presented the results regarding IADE's three CALL qualities: Meaning Focus, Learner Fit, and Impact. The same types of quantitative and qualitative data were analyzed. Survey data obtained from participants' responses to Likert-scale, yes/no, and open-ended questions revealed their perceptions, and the transcripts of think-aloud protocols, observations, and semi-structured interviews not only helped explain learners' viewpoints,

but also understand the nature, causes, and consequences of their interaction with different features of IADE and its automated feedback.

Evidence gathered for all of the three qualities was predominantly strong. Although at different points in the revision process, most of the participants did focus on discourse meaning, and such focus appeared to be prompted by the program's color-coded feedback and facilitated by its Help Options. Displaying colors for moves other than intended by the learners, it made them notice that the expressed functional meaning differed from what they had in mind. This triggered their reflection on functions of moves, during which learners generally discovered a connection between certain functions and linguistic means, and eventually lead to construction of new functional meaning. IADE's feedback also seemed to be a determining factor for the type of learning experience the participants had, color-coded feedback having a more positive impact than the numerical feedback. The former, by fostering focus on meaning, seemed to stimulate deeper cognitive involvement, more effective revision strategies, motivation and more positive emotions. The latter, on the contrary, could be limiting in its guidance through percentages and comments on the proximity of the draft to disciplinary writing norms, leading to learner's behavior that was likely to negatively affect them at different levels. However, the participants evaluated various characteristics of IADE's feedback mostly as excellent and good. It was regarded as appropriate and helpful because it was individual and discipline-specific, output-focused, iterative, and adequately negative. Also, the availably of Help Options was important for successful completion of the revision task.

The next chapter draws conclusions about the effectiveness of IADE and its automated feedback for formative assessment purposes. It makes connections among

different evidence gained in this study, ponders over limitations, and discusses the implications of the obtained results for instructional implementations of IADE in particular and of automated writing evaluation in general. Given that this dissertation research draws from different theoretical perspectives, implications in view of the utility of those perspectives are also discussed. The chapter concludes with recommendations for future work.

Chapter 7. Conclusion

The purpose of this mixed-methods study was to evaluate the overall effectiveness of IADE and its automated feedback in the formative assessment context of NNS graduate academic writing instruction. This final chapter summarizes the key findings based on quantitative and qualitative evidence and outlines the conclusions that were drawn regarding IADE's CALL qualities, which framed the evaluation. It is argued here that the effectiveness of IADE is a result of combined strengths of its Language Learning Potential, Meaning Focus, Learner Fit, and Impact. These qualities, which are closely related and interdependent, are enhanced by the automated feedback of IADE, making this program an effective formative assessment tool suitable for implementation in the targeted instructional context. After the main results are reiterated, the limitations of the study are discussed, explaining the degree of generalizability. Then, the implications for academic writing pedagogy, formative assessment, AWE implementation, and theoretical knowledge are considered. Finally, recommendations for future research are made and summative conclusions are drawn.

7.1. Research questions answered

7.1.1. Language Learning Potential

The Language Learning Potential quality was the most essential and the most informative in evaluating the effectiveness of IADE and its automated feedback. Two research questions were posed: (1) What evidence suggests that the feedback provided by IADE leads to students' noticing of and focus on discourse form? and (2) What evidence suggests that students learned the target discourse forms that were focused on during the

interaction with IADE? To answer these questions, various quantitative and qualitative data were collected and analyzed. Quantitative data consisted of Likert-scale, yes/no, and open-ended survey responses; automated and human scores for first and last drafts; pre-/post test scores; and frequency counts for draft submission and access to Help Options. Qualitative data contained 105 students' drafts and transcripts of think-aloud protocols and Camtasia recordings, observations, and semi-structured interviews provided by the same sample of 16 participants. Overall, all these data helped answer the questions about IADE's Language Learning Potential quality and exhibited evidence supporting its strength in terms of five aspects: noticing and focus on discourse form, improvement in the rhetorical quality of writing, learning gains, practice, and modified interaction. Some data sources, however, pointed to less encouraging phenomena.

The first positive claim is that focus on discourse form occurred, being triggered by both numerical and color-coded feedback as well as by IADE's Help Options (accessed with an average of 12 clicks per student), which helped the learners notice less salient characteristics of rhetorical moves, particularly those that were expressed with move-specific vocabulary. Introspective data suggested that focus on form was the head of a learning cycle, whose elements were negative evidence, enhanced understanding of moves, and output modification. These elements occurred causatively and sequentially with every iteration of the cycle, which was prompted by the automated feedback.

The second positive claim – that IADE contributed to improvement in the rhetorical quality of writing – found support in participants' perceptions of how and why improvement occurred. It is also based on evidence of improvement in participants' work products, which was statistically significant according t-test results on both automated and human scores for

first and last drafts. Manual analysis of a sample of 32 first and last drafts not only showed that the rhetorical quality indeed improved; it also revealed that the modifications leading to improvement were made to content, structure, vocabulary, grammar, and mechanics.

Learning gains were also statistically significant as demonstrated by t-test results; after having revised their work products with IADE, the students enhanced their knowledge of moves and steps. Also, participants' judgments of causes of learning and of observed improvement mirrored the introspective evidence regarding the elements of the discovered learning cycle, supporting the strength of IADE's Language Learning Potential quality.

The roles of such aspects as practice through multiple resubmission and modified interaction through IADE's Help Options did not find strong support in the results of inferential statistical analyses. Although most of the participants thought that improvement was facilitated by both resubmission practice and modified interaction, Pearson product-moment correlation analyses revealed a statistically significant relation between practice and overall draft improvement, but not between modified interaction and overall improvement. This could be a small sample size effect. Alternatively, the degree of newly acquired declarative knowledge or other individual learner characteristics may have also been a reasonable explanation for the lack of a significant relationship. Further investigations would be needed to substantiate these hypotheses. In this study, however, multiple qualitative data showed that the participants engaged in modified interaction with different types of IADE's Help Options for different purposes such as knowledge consolidation, search of examples, comparison with disciplinary writing, and strategy development.

7.1.2. Meaning Focus

In relation to Meaning Focus, the research question posed was: What evidence suggests that IADE's feedback directs students' attention toward the meaning of the targeted discourse forms? Survey questions as well as the think-aloud protocols and Camtasia recordings, observations, and semi-structured interviews were analyzed and yielded mostly positive evidence for two aspects of this CALL quality: focus on functional meaning of discourse and construction of such meaning. When triangulated, all the data sources pointed to the same phenomena and helped gain an understanding of the causes and consequences of Meaning Focus.

It appeared that focus on the functional meaning of the discourse occurred primarily when the color-coded feedback displayed a mismatch between intended and expressed functional meaning, which makes this particular form of feedback most potent for stimulating learners' focus on meaning. Noticing that there was a mismatch, learners engaged in reflections such as self-questioning, self-verification, and self-planning or took actions such as output modification and consulting IADE's Help Options. Gradually, they discovered a connection between functional meaning and vocabulary choices, which helped them construct new meaning more effectively.

Additionally, causative effects were observed: noticing a mismatch between intended and expressed meaning led to reflection on functional meaning, and making connections between functional meaning and lexical choice led to construction of new discourse meaning. The latter observation, however, must be viewed with caution. On the one hand, it was quite encouraging to find that the students became cognizant of the fact that certain functional meaning can be rendered through certain lexical items; this helped them identify move-

specific phraseology and use it to successfully modify their output. On the other hand, it was discouraging to see that, once they realized this connection between functional meaning and vocabulary, they tended to rely more on lexical modifications, making this their main revision strategy. Such a consequence is a potential weakness of IADE's Meaning Focus quality.

7.1.3. Learner Fit

The research question for the third CALL quality, Learner Fit, was: What evidence suggests that the formative feedback on discourse forms is appropriate for students with the characteristics of targeted learners? Positive evidence was found in survey responses and in the think-aloud protocols, observations, and semi-structured interview transcripts, indicating that the program and its feedback were perceived by learners as appropriate for a number of reasons.

The learners perceived IADE's feedback as comprehensible, appropriate, and helpful. It was also motivating in that it stimulated their intrinsic motivation to follow disciplinary norms of their academic community and that it approximated computer interaction to interaction with human professionals. Of nine feedback characteristics, the participants evaluated the color-codes as most helpful, also highly appreciating iterative, output-focused, individual-specific, negative, intelligent, and metalinguistic features of the feedback. The introspective results indicated that the feedback was suitable for individual learner characteristics and appropriate for the completion of the revision task. In addition, all qualitative data suggested that the program's non-prescriptive way to provide feedback can

support learner control and that the availability of the Help Options can facilitate their problem-solving by encouraging modified interaction.

Shortness (defined as feedback is provided concisely in that it briefly presents the descriptive percentages representing the use of moves) and explicitness (defined as feedback is demonstrated clearly and fully) were viewed as weak characteristics, which could have been stronger had the program the ability to generate feedback not only on moves, but also on steps and on language errors and had it the ability to provide specific directions for remediation. Also, even though feedback comparing students' drafts to the norms in their discipline was positively appraised by most of the participants, the few students whose particular discipline was not represented in IADE could not benefit from it.

7.1.4. Impact

The last research question answered was related to Impact: What evidence suggests that students have a positive learning experience with IADE? Both positive and negative evidence was found in participants' responses to Impact survey questions and in the transcripts of the think-aloud protocols/Camtasia, observations, and semi-structured interview recordings. All these data showed that that the participants were influenced at different levels: affective, intrinsic, pragmatic, and cognitive, which seemed to have had certain effects on one another.

Both types of IADE's feedback, numerical and color-coded, were perceived as motivating and beneficial by the participants. However, they appeared to have opposite effects on learners' revision process. The impact was likely to be negative when learners relied on the numerical feedback; in this case their cognitive involvement was low, which

lead to ineffective revision strategies and sporadic output modifications and, therefore, resulted in frustration and disappointment. Positive impact, on the other hand, occurred when the learners focused on the color-coded feedback, in which case the degree of cognitive involvement increased. Therefore, they employed more effective revision strategies and made consistent and more successful changes to their drafts. Consequently, the learners experienced positive affective impact, feeling excitement about their accomplishment and motivation to continue revising.

Negative Impact was common at the beginning of the revision process; however, the learning experience curve changed its direction towards positive Impact when the learners reached a turning point. It was the point when the color-coded feedback drew their attention to functional meaning, which triggered deeper cognitive involvement that was decisive for positive effects at pragmatic, affective, and intrinsic motivation levels.

7. 2. Interdependency of IADE's CALL qualities

Chapters 5 and 6 elaborated on the strength of IADE's CALL qualities individually; nevertheless, much of the evidence obtained reveals ways in which they are interrelated. As summarized in Figure 7.1, the Language Learning Potential quality, whose core was a learning cycle with four iterative elements, appeared to be dependent on the strength of Meaning Focus, Learner Fit, and Impact. Meaning Focus seemed to be crucial for positive Impact, which also depended on the degree of Learner Fit. Therefore, the findings about IADE's overall effectiveness should not be interpreted separately.

Notable intersections could be distinguished between Language Learning Potential and Meaning Focus, more specifically, between the learners' process of focus on functional

meaning and the elements of the learning cycle discovered in the analysis of the Language Learning Potential. Noticing of negative evidence was realized in Meaning Focus through learners' noticing of the mismatch between intended and expressed meaning. Enhanced understanding occurred when they identified a connection between lexical choice and functional meaning. Lexical changes were a form of output modifications. Consequently, it may be inferred that focus on functional meaning contributed to learning and improvement.

Learner Fit is also related to the learning cycle in Language Learning Potential due to IADE's feedback and Help Options. The individual-specific and output-focused feedback enhanced the Language Learning Potential quality by making the learners focus on discourse form and notice negative evidence in their writing. The Help Options, especially the annotated corpus, also reinforced focus on form and encouraged modified interaction, which contributed to learners' understanding of targeted rhetorical conventions.

Finally, positive Impact appeared to spring from Meaning Focus and to eventually influence the Language Learning Potential. More specifically, positive Impact began when learners realized the role of functional meaning, which reinforced the re-occurrence of the learning cycle. On the one hand, Meaning Focus caused deeper cognitive involvement that lead to enhanced understanding; on the other hand, it stimulated the development of effective revision strategies that led to effective output modifications. Moreover, positive affect increased learners' intrinsic motivation to go through the learning cycle repeatedly during their revision with IADE and to improve their written discourse. Considering these intersections, and considering that the findings of this research revealed that each of these individual CALL qualities is relatively strong, it can be concluded that IADE is an appropriate AWE application for the given context.

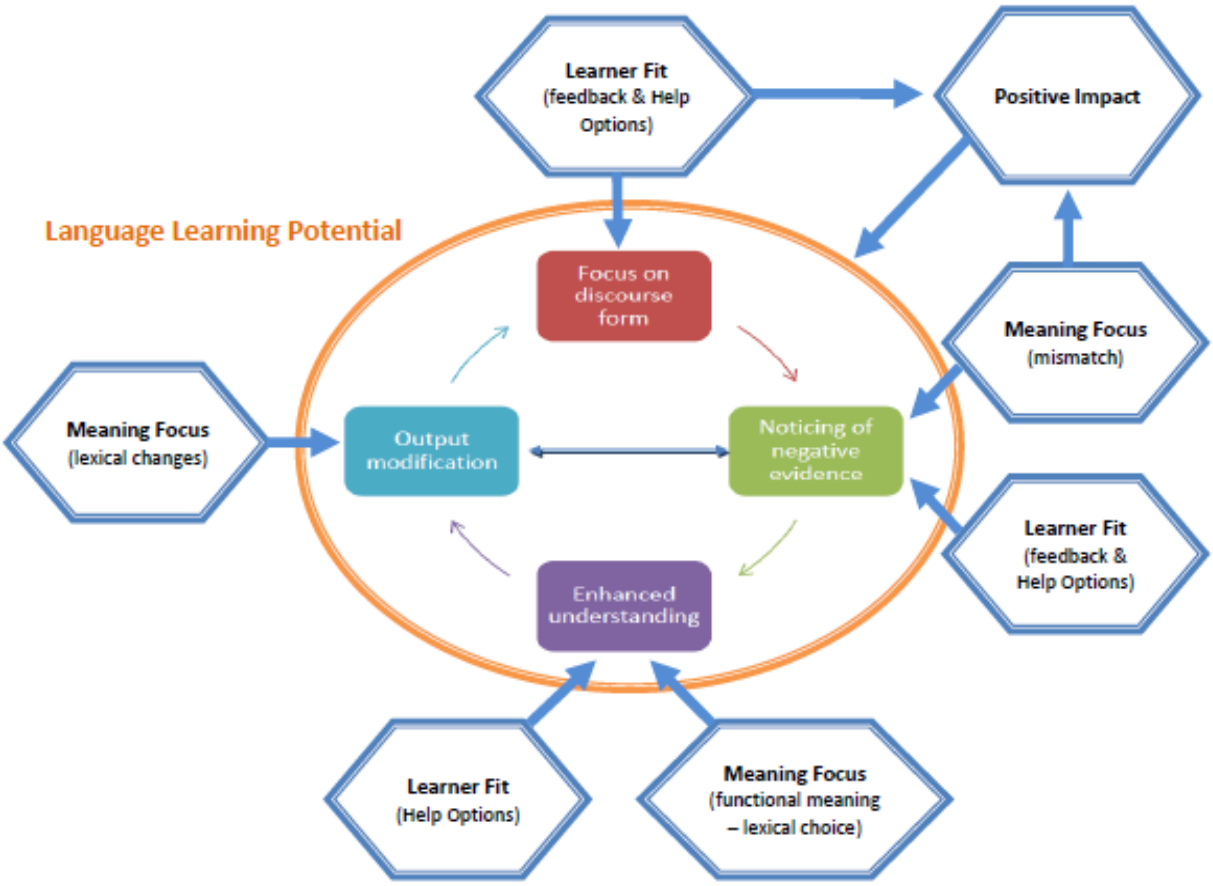


Figure 7.1. CALL qualities interrelated

It is of utmost importance to emphasize that the implementation context is a key factor that can determine the strength of the CALL qualities, and, therefore, this conclusion may be generalizable for implementation contexts that are the same or similar to the instructional context of this study, for example, at universities in the USA and in other English-speaking countries that enroll international graduate students. A similar context would have to share a number of critical aspects that made IADE work the way it did: the teaching of the research article genre, the specific research writing goals of the course targeted to an academically heterogeneous group of students, the characteristics and the level

of language ability according to which the students were placed in this course by the institutional English placement test, the corpus-based pedagogy, the use of the WordSearch and Callisto tools, the beliefs of the instructors who employed corpora in the teaching of academic writing, and the computer-based classroom environment. IADE, or a similar program, would have less learning potential and, therefore, would be less effective if one or more of the other CALL qualities were weak or faulty due to context differences, e. g, degree of integration, instructional goals, pedagogical approaches and strategies, targeted learners, etc. Undoubtedly, IADE would be of no value for non-academic writing instruction or for undergraduate level composition courses. Also, even though the Help Options contain instructional materials and annotated corpora, by no means is IADE meant to replace classroom instruction. As demonstrated in this dissertation, it was effective when implemented as a supplemental formative tool used for learning a subset of course objectives.

7.3. Limitations

Inasmuch as the results are encouraging, there were limitations that restrict their generalizability. These limitations are related to the research design and to some aspects of data collection tools and analyses. The fact that the study was conducted in an instructional setting, in which students take writing courses in order to complete language proficiency requirements enforced by the university, made it impossible to employ a rigorous methodology that would rely on random sampling and assignment of participants from the population of interest. All the participants were subject to the same type of treatment because the alternative of experimental and control groups was not possible or desirable due to a potential fairness issue. Dividing the participants into control and experimental groups would

have exposed them to different instructional approaches and learning opportunities, and that might have affected their grades in this academic writing course, which were relatively high stakes for the students because they counted towards the grade point average in the graduate program. However, this design is somewhat disadvantageous. Examining work products of learners who had and had not used IADE, not only for score comparison, but also for the nature of the revision process and of the quality of final drafts, would have allowed for a better understanding of how learning to write academically with and without such an AWE application occurs.

Another limitation is that the introspective data were collected only during one-time sessions, which were held in the presence of the researcher. Some findings would perhaps be expanded or altered if data from the first to the last draft submissions were available and if the interaction with the program were tracked in less supervised conditions, for instance, if the participants were recorded on video. It is possible that the video recordings would have offered more in-depth perceptions and more genuine reactions. If video recording was done, direct observations could still be possible since the technology could be set for synchronous viewing in a remote location. Additionally, considering how informative the immediate post-revision interviews were, more insights could have been gained through delayed post-revision interviews after the final drafts were submitted to the instructor.

All the data collection instruments used in this study yielded appropriate and sufficient evidence. The only tool that proved to be somewhat less effective than expected was the survey. Perhaps because it was too long, some participants did not answer some open-ended questions towards the end of the survey. Also, in some cases, they provided inappropriate answers to “why” and “how” questions accompanying a yes/no question. The

limitation is not only in the instrument; no clarification follow-up was initiated, and therefore not all students' views were captured and understood.

Finally, although the dual researcher/teacher role is not uncommon in classroom-based research, the fact that the researcher was also the instructor of four out of six groups of participants, among which were the 16 students from whom think-aloud, observations, and interview data were collected, may be viewed as a limiting factor as well. The teacher's background, beliefs, relationship with students, may have influenced the nature of the data. However, if there were outcomes that should be attributed to teaching, they would be related to the stage of classroom instruction and the acquisition of declarative knowledge, and not so much to improvement outcomes since interaction with IADE was only introduced in class and took place mostly out of class.

7.4. Implications and recommendations

The findings of this dissertation have implications for both practical and theoretical perspectives that informed it. Academic writing, formative assessment, and AWE are the areas related to practice. The implications for academic writing are discussed in view of the potential benefits to both students and teachers, implementation recommendations being made as well. The implications for formative assessment are pondered over in view of how IADE supported the goals of formative assessment and what qualities of effective formative feedback it appeared to possess. With respect to AWE, parallels are made between certain results of this study and some claims related to the AWE debate, the main implication being that the research methodology employed here may be a viable model for further AWE classroom research as it allows for gathering appropriate empirical knowledge to address

existing disagreements. Then, the SFL, SAT, and SLA theories that proved to be very informative for this work are discussed in terms of their utility for practice.

7.4.1. Implications for practice

- Academic Writing

IADE caters to the needs of both teachers and students. It can help teachers compensate for their lack of knowledge of the rhetorical composition of disciplinary texts. In addition, it can assist teachers in providing formative feedback to their students, allowing them to devote more time to other pedagogical matters that may require their efforts. In terms of pedagogy, IADE takes the corpus-based approach currently employed in instruction to the next level by making the corpus not only a concordancing resource used for observation and discovery in class, but also an annotated resource used for application and experimentation during independent revision.

The empirical insights gained in this study can also benefit the teachers, helping them better understand why and how different theoretical perspectives can and should guide writing instruction, and more importantly, how interacting with a program like IADE can help their students. First, IADE can offer sufficient practice opportunities and instant individualized guidance. Second, it can help students gain a better understanding of academic discourse conventions and improve the quality of their writing. Third, it can support and reinforce learners' focus on functional meaning, making them realize how the context defines the meaning potential of language and the range of language choices that can be used for particular purposes. Being able to understand and construct functional meaning can replace mere replication of prescribed norms with conscious discourse choices. Finally, with its

iterative feedback and the availability of Help Options, IADE can scaffold and facilitate revision, which is crucial for the writing process.

With respect to implementation, a number of recommendations must be made. Both the teacher and the students should be well acquainted with the functionality of IADE and its potential benefits and caveats. The teacher should introduce the program to the students only after they are taught Swale's move schema and after they consolidate the new knowledge by completing corpus exploration activities in class. Insufficient declarative knowledge may influence students' revision process with IADE; it may become lengthier and more cumbersome since the students would be very likely to spend much time on trying to understand the functions of moves and steps instead of focusing on what can be done to improve them. Also, the teacher should allow sufficient time for the revision task. An important finding of this study was that learners' revision process involved both weaker and stronger strategies before and after they developed a focus on functional meaning; therefore, the students need enough time to be able to reach the meaning focus turning point after which they begin making more successful output modifications.

- Formative assessment

A particular aspect of NNS academic writing instruction that IADE was meant to enhance was formative assessment. The findings of this dissertation allow for the claim that IADE's automated feedback can be considered an appropriate formative means for the given context. Various evidence obtained here helped understand how IADE enhanced the overall purpose of formative assessment and which of the qualities of formative assessment discussed in Chapter 2 it exhibited.

According to McMillan (2007), the major objectives of formative assessment provided during instruction are to improve student motivation and learning. The evaluative results on IADE's CALL qualities suggest that it was able to fulfill this goal. Learner Fit and Impact findings revealed that the program's automated feedback was a powerful motivator because it stimulated the learners' to follow disciplinary academic writing norms and to engage in careful revision, devoting greater efforts to progress from draft to draft. As for learning, the results on Language Learning Potential demonstrated not only that learning and improvement occurred, but also that the feedback played an important role for these processes to unfold.

The program's strong Meaning Focus was a key factor that induced cognitive involvement, which is one of the most important qualities of formative feedback as it is believed in educational psychology. The color-coded feedback triggered learners' cognitive mechanisms by making negative evidence salient and thus pointing to gaps between actual and intended discourse meaning. Another essential quality of formative feedback is goal orientation. Learners' willingness to work with IADE in order to improve their academic writing skills, to pursue the challenging revision task, and to employ new strategies for the completion of that task speaks of the learning type of goal orientation, which is most desirable for learning (Fisher & Ford, 1998). The appropriateness of specificity and timing qualities of formative feedback were supported by Learner Fit results, which also suggested that IADE's feedback played directive, informative, advising, and summarizing roles that are thought of as characteristic of effective formative feedback (Sales, 1993). The only weaker aspect of this program's feedback was complexity; the informative tutoring nature of the

feedback (see Narciss & Huth, 2004) was not as potent as the learners wished it would have been. While the automated feedback was sufficiently verifying, it was not elaborative enough; in other words, it did not provide sufficient direction for remediation. A preemptive attempt to address this drawback was the inclusion of guiding revision tips in the program's Help Options.

- AWE

Pedagogical uses of AWE have been debated in the literature largely due to judgmental speculations that lack empirical support. While this study does not prove the effectiveness of AWE applications across different instructional contexts, it provides a viable methodological model for future AWE evaluative research as it allows for the generation of more complex and more comprehensive results. This model can be very appropriate for future validations of AWE systems as it allows capturing the complexities of both the processes and the products resulting from AWE use. Evidence yielded from such research can be used to build an empirical evaluation argument for the use of this technology.

In this study, for instance, in addition to yielding evidence that corroborate outcome-based research, according to which students' writing improves as a result of interaction with an AWE application (Attali, 2004; Elliot & Mukulas, 2004; Foltz, Laham, & Landauer, 1999; Leah Rock, 2007), this model made it possible to explain why and how improvement occurred. Also, the findings not only showed that learners were satisfied or dissatisfied with the software as in Chen and Cheng (2008), Elliot and Mikulas (2004), and Yang (2004); they revealed when and why positive and negative learning experience occurred. Additionally, the Meaning Focus category allowed for obtaining evidence that AWE, if designed and

implemented appropriately, does not present the danger of undermining learning or subordinating meaning, as feared by Cheville (2004). On the contrary, as in the case of IADE, it may be capable of directing learners' attention towards meaning, which is one of the major factors that enhance learning to write.

7.4.2. Utility of theoretical approaches

- Systemic Functional Linguistics (SFL)

The SFL perspective helped identify a major weakness in the pedagogy of graduate-level academic writing courses at ISU – the lack of a focus on language as a systematic resource, which the students need to be able to use in order to build a discourse whose rhetorical structure would convey multifunctionality and higher-level semiotics. To address this weakness, the IADE program was conceptualized so that functional meaning was integrated in its automated analysis of rhetorical moves. Being trained on texts annotated for moves and steps, the program can recognize and provide feedback on these elements. The functionality of its Help Options also relies on annotated corpora, where each sentence is glossed with its function.

The findings allow for the conclusion that, without IADE's color-coded feedback and input enhancement, which were the core means of drawing learners' focus to functional meaning, their final work products would have been less effective. Mere identification of structural components of texts, observed by learners in their disciplinary corpora in class, and attempting to replicate model structures would not have been sufficient for them to develop the skill of using appropriate rhetorical and linguistic conventions. When learners focused only on the discourse form reflected in the distribution of their moves, they did not produce

effective discourse elements that would clearly define the genre to which their texts were meant to belong. Those elements began to appear only when the learners realized that texts are made of functionally organized language, which encodes various meanings. For instance, the introspective insights revealed that it was focus on functional meaning that played a key role in the rhetorical improvement of learners' drafts. Although indirectly, these results suggest that language functions and semantics are indeed central to communicative activity, supporting the informative value of SFL for the teaching of academic writing.

- Skill Acquisition Theory (SAT)

Guided by the SAT, IADE was meant to account for the procedural, or practice, stage of skill development. It was intended to provide learners with practice opportunities by giving them the option of resubmitting their drafts as frequently as needed. Although a relation between practice and improvement was expected and qualitatively observed, that relation was confirmed statistically only for overall draft improvement and not for the improvement of individual moves and length. This finding may imply that multiple re-submission may not have been a sufficient condition for learners to transform their new declarative knowledge into automatic behavior. It is possible that, if IADE were used for practicing the writing of individual moves separately, with more time being given for each move, the relation between practice and improvement might have been stronger. Another possible explanation may be the fact that the participants worked on the revision of a single text, and, to be able to reach the production/automatization stage of skill development, they might need to practice writing more texts, applying their knowledge of rhetorical conventions to different contexts and topics. Nonetheless, the overall results did suggest that improvement

occurred and that practice through resubmission was perceived by learners as playing a fundamental role for improvement.

Another relevant point is that SAT postulates that practice is essential for gradual changes in the cognitive mechanisms activated by learners to carry out a certain task. There was evidence in this study that practicing with IADE triggered cognitive mechanisms, which resulted in enhanced understanding and output modifications that were repeatedly submitted, and that, although tangentially, also adds substantiation for the assertions of this theory.

- SLA interactionist approach

The interactionist perspective similarly received indirect support for the role of its constructs as applied to computer-assisted learning, which can be inferred from the qualitative insights about the Language Learning Potential cycle. Learners' modified input, which was provided to them in the form of their own automatically processed output, became a productive source for their linguistic hypotheses. Feedback stimulated computer-learner interaction by confirming or invalidating those hypotheses, prompting revised output. What became particularly prominent in this study was the value of noticing of negative evidence. It lead to intra-personal interaction during which learners appeared to construct a better understanding of discourse phenomena and to eventually produce better quality output. It was also interesting to see how potent input enhancement was. The color-codes made noticing of negative evidence more salient and therefore stimulated learners to test hypotheses, which they generated with regards to the nature of the miscommunicated message.

The main implication is that theoretical interactionist models can be more confidently extrapolated to computer assisted language learning environments, in general, and to NNS academic writing, in particular. On the one hand, future studies investigating learners' interaction with AWE and ICALL applications can add support to SLA theoretical claims and strengthen them with new empirical evidence. On the other hand, developers of these new learning technologies can rely on interactionist models in designing and evaluating their products, which, while highly complex, are not always theoretically informed.

7.5. Directions for future research

It is hoped that IADE, being a new, unique AWE tool, will arouse the interest of Applied Linguists. Replication studies conducted with it in other similar contexts will help determine which phenomena are generalizable across contexts and which are strictly context-dependent. They will help clarify how context factors can influence the strength of this program's CALL qualities and whether that would change the nature of the relations between and among them and the degree to which one would influence another.

Other future research can extend the investigation of individual CALL qualities by focusing on some aspects of the findings that present interest as stand-alone topics. For instance, the value of practice opportunities for Language Learning Potential was discernible in the qualitative data, but did not receive sufficient quantitative evidence. A new study can not only look into whether the number of revisions correlates with higher final scores; it can inquire about what the optimum number of revisions per draft would be. It can also examine whether the types of revisions correlate with score improvement and whether this

relationship is mediated by other factors such as language background, proficiency level, learning style, and revision strategies.

Another study can concentrate on Meaning Focus. It has been claimed in this dissertation that the learners were able to construct meaning for specific communicative purposes partially because they were able to identify a connection between functional meaning and formulaic expressions. However, bearing in mind that the program's analyzer was developed based on a lexical approach, at some point, learners may realize that IADE's analysis is vocabulary-based and, therefore, adjust their writing and limit their revisions to lexical changes. Further research is needed to clarify whether this would indeed be an unavoidable consequence. The findings would also have direct implications for future uses of the n-gram NLP approach in the development of other writing applications.

The Learner Fit results revealed why and how IADE was appropriate for learners. The next step in this respect can be to investigate whether it is appropriate enough to create an environment where the learners can develop into reflective writers with a sense of purpose for their academic audience as well as with a sense of identity that relates to that audience. Findings in this vein would shed light on whether and how this program in particular, and AWE in general, may support social and communicative dimensions of writing.

In terms of Impact, claims have been made that positive or negative learning experience appears to be dependent on the type of feedback and the revision strategies employed before and after the turning point of focus on meaning. Future research can zoom in on what this process implies, what it takes to reach this turning point, whether it is possible that this point is not reached, and what would be the consequences of that. It is important that

this process is explained theoretically and that recommendations on how to best address this phenomenon pedagogically are put forth.

Moreover, the Authenticity and Practicality CALL qualities outlined in Chapelle's (2001) conceptual framework, but not explored in this dissertation, need to be pursued in future work to strengthen the claims made about IADE's effectiveness. An investigation of Authenticity should examine whether students' performance on the task completed with IADE corresponds to what is expected of them outside the classroom. Evidence on Practicality, in turn, should help determine whether the resources used to create and support the program and its use are adequate.

Another future avenue is to expand the evaluation of IADE's effectiveness to include the perspectives of stakeholders other than learners, e.g. teachers and administrators. This would bring into the picture the broader ecology of AWE use in an instructional context. Would the findings uncover that this technology, relieving teachers of the burden of commenting on multiple individual drafts, can be beneficial in other ways as well, or would it contribute to depersonalizing the classroom environment? Would its implementation positively or negatively influence the broader goals of teaching and learning? Answers to such questions may be able to reconcile some opposing views in the AWE debate.

Furthermore, the results of this dissertation provide a foundation for future AWE work under the umbrella of formative assessment. The formative capabilities of IADE's feedback help realize that the areas of formative assessment and AWE can benefit each other - formative assessment can be enhanced by automatization and individualization of feedback, while AWE can find valid and legitimate uses for its technological potential in teaching and learning contexts. Researchers may find inspirations to propose specific methods of

investigating the formative potential of automated feedback – how objectively and reliably it can point to 'gaps', whether it can serve as proper scaffolding, to what degree it can allow learners to monitor quality during writing and revision, whether it can enhance goal-orientation by offering opportunities to self-assess progress, etc. Widely marketed AWE applications like Criterion, MyAccess, and WriteToLearn possess complex feedback capabilities and can serve as excellent platforms for such research.

Finally, a further development path is to expand IADE to include other research article sections in addition to Introductions. In fact, such work, funded by the Computation Advisory Committee at Iowa State University, has already begun and is now at the stage of annotating Discussion and Conclusion sections.

7.6. Conclusion

To conclude, this dissertation makes a considerable contribution to the field of Applied Linguistics. It introduces IADE, an innovative exemplar of learning technologies worth replicating not only because of its demonstrated effectiveness and appropriateness, but also because of its theoretically and empirically grounded conceptualization and design. More importantly, by investigating IADE's Language Learning Potential, Meaning Focus, Learner Fit, and Impact qualities, it provides valuable empirical knowledge to the areas of NNS academic writing, formative assessment, AWE, and ICALL. Despite a few methodological limitations, which were unavoidable in the context of this study, current results have a number of important practical and theoretical implications and are informative both for future research and applications of this technology.

Appendix A. Pre- and post tests

Pre-test:

Task I

Using Swales' Move schema, identify the move and the step for each sentence provided below and write them in parentheses after each sentence.

Example: The remainder of this paper is divided into five sections. (Move 3, outlining the structure of the paper)

1. That is, regardless of whether exposure to the disc is assumed to have a largely positive or negative impact, we would expect the perceived effects on oneself would be more toward the positive end of a scale ranging from positive effects to no effects to negative effects.
2. There is much current interest among space physicists in the transport, acceleration, and loss of relativistic electrons in the outer radiation zone of the Earth's magnetosphere.
3. In our experiments casein did not inhibit, but in some cases cow's milk enhanced, the absorption of supplemental iron at the traditional level of fortification.
4. An exhaustive survey of the literature reveals that no work has been reported on the reactions of different nitriles with 16-hydroxyhexadec-cis-9-enoic acid in strong sulfuric acid medium.
5. Note that in the remainder of this paper, we term Object Database Management Systems (ODBMSs) both object-oriented and object-relational systems.
6. The results of this basic investigation can help to understand the unsteady complex flows occurring in turbomachinery.
7. In this study, the direction vector was obtained by using the Z-map method and the production of this vector was based on the fact that the bottom part of the mill- ing cutter rode on free-formed surfaces.
8. The understanding of the physics of such a separation bubble under the influence of incoming wakes is crucial to determine the profile losses and aerodynamic performance of the blades.
9. We set to find whether the social construction of knowledge occurs in both forums and whether the participation protocols affect knowledge construction and participation.
10. This suggests that a study of this topic is timely and may provide the basis for future revisions in currently accepted correlations.
11. Today's state of the art turbine blades, so-called controlled diffusion airfoils, allow a small separation bubble on the blade surface.
12. In equilibrium, cross-sectional variation in the benefits and costs associated with gathering and disclosing information leads to variation in information asymmetry.
13. The following sections of this paper describe the background to this problem and review the relevant literature; outline the research approach; report on learning from case studies; report on learning from interviews; describe the questionnaire survey

used to establish the preferred tool format; describe the development and testing of the checklist; and discuss the findings and draw some conclusions.

14. If this tendency were accounted for when estimating efficiency, results could differ substantially.
15. Redner and Walker (1984) pointed out that the ML approach for finite mixture model could encounter unbounded likelihood in some special cases.
16. Hathaway (1985) suggested that using simple constraints in an optimization problem can lead to a strongly consistent and global solution.
17. Still, can it be claimed that this method can reduce spherical aberration to acceptable levels?

Task II

Annotate each sentence of this research article introduction in terms of moves and steps. Do this by using the 'Insert comment' function of Word.

Multinational corporations rank the top 10 global communications challenges and offer some valuable potential solutions. International data communication (IDC) activities are vital to multinational corporations (MNCs) in managing the complexity of information exchanges required for the control and implementation of worldwide business strategies. IDC activities are also critical because they allow MNCs to extract information and move it globally without incurring major time delays, or transportation, reproduction, or inventory costs. Compared with data communications that occur within national borders, IDC encounters more difficulties due to differences in the international technological infrastructure, network systems, culture, government regulations, and the level of technological and economic development. Surprisingly, there has been little empirical research to identify the issues MNCs encounter when performing IDC, and needless to say, no solutions have been widely reported to deal with these issues effectively. However, several empirical studies have been conducted on the issues of information systems management. One of the earliest studies (1982) surveyed members of the Society for Management Information Systems (SMIS). Four subsequent studies, based upon each other, were performed on the SMIS membership since then in an attempt to build a cumulative research tradition of IS issues in the U.S. Similar studies have also been conducted in other countries to essentially identify and rate the relative importance of domestic and international IS (IIS) issues as implied by current competitive priorities, major concerns, and technology advancements. Although these studies do not produce uniform findings, network-related concerns have consistently been rated as top IS issues. This article reports on a two-round Delphi study to prioritize the IDC issues confronting MNCs. A questionnaire was administered to IS executives at 300 Fortune 500 companies. Only U.S.-based MNCs were selected because we wanted to improve the study's response rate by avoiding international mailings. The IIS background of the responding MNCs was evaluated and unqualified respondents (those with little or no IIS experience) were removed throughout both rounds. The results were used as the basis for the IDC issues study. However, we also conducted follow-up interviews with 11 participants to collect more in-depth responses to the IDC issues. Our data suggested that these issues may have serious implications, which we discuss in great detail in this paper.

Post-test:

Task I

Using Swales' Move schema, identify the move and the step for each sentence provided below and write them in parentheses after each sentence.

Example: The remainder of this paper is divided into five sections. (Move 3, outlining the structure of the paper)

1. The characterization and modeling of gene-regulatory networks have been pursuited since the early 1970s.
2. It utilizes aqueous ammonia in a recycle mode, thus termed as Ammonia Recycled Percolation (ARP).
3. In this study, we explored how the cell can locally transduce the rigidity of the ECM into a contractile signal used to direct cell edge extension.
4. However, to our knowledge, only two surveys have addressed the extent to which UML diagrams are used in practice (Grossman, Aronson, & McCarthy, 2005; Zeichick, 2002), and neither examined why analysts choose to use some diagrams and ignore others.
5. A typical representative of these polymers is polyimide w6,7x. (Move 1, Making topic generalization)
6. Because of their potential impact on the macroeconomy, stock market movements are likely to be an important determinant of monetary policy decisions.
7. Given the rising instability of marriage during this period, an incentives theory must focus on the probability of staying married as well as the probability of getting married.
8. Here, the synoptic patterns associated with rain events in the vicinity of the jet are compared to synoptic patterns in the context of rainfall in the SACZ, and vice versa.
9. The results suggest that the main determinants of survival and growth were indeed mill size and mill age.
10. The present paper is organized as follows: Section 2 provides a brief description of the four galaxies in the sample; Section 3 describes the observations and the data reduction, with emphasis on the derivation of emission-line images; Section 4 presents the derivation of the line ratio images; Section 5 describes how nonphotoionized gas is discriminated from photoionized gas and presents the main findings of this work.
11. While the main advantage of the RP technology is that the manufacturing process is independent of the geometric complexity of the part, RP has its own drawbacks, which include poor surface finish and low efficiency.
12. The goals of the study reported here were to investigate the levels of mRNA for select proinflammatory cytokines to begin to obtain a further understanding of the pathogenesis of the disease induced by these problem pathogens.
13. To what extent is the language learner's ability to focus on form related to literacy development?
14. It had been proved that the PRRS virus, an arterivirus of swine, which is believed to

play a key role in the porcine respiratory disease complex [14] and [15].

15. This study contributes to the literature in a number of ways.
16. We introduce new tests lending support to the hypothesis that the discontinuities in earnings distributions are associated with accrualsbased earnings management in our sample.
17. Still, can it be claimed that this method can reduce spherical aberration to acceptable levels?

Task II

Annotate each sentence of this research article introduction in terms of moves and steps. Do this by using the 'Insert comment' function of Word.

For individual taxpayers, dividend income is taxed at a higher effective rate than capital gains income. As a result, there is a tax penalty to receiving dividend income for individual taxpayers. This penalty is often referred to as a dividend tax penalty. An actively addressed question in the economics, accounting, and finance literatures is whether a dividend tax penalty is capitalized into the return on a firm's common stock. Studies that examine this question are important because their findings can shed light on fundamental issues such as the cost of equity capital and the optimal dividend policy. In summarizing the research in this area, Fama and French's (1998) assessment is that no consensus has yet emerged on the question of whether a dividend tax penalty is capitalized into the return on a firm's common stock. The purpose of this paper is to provide additional evidence on this question. To summarize, the evidence that we provide suggests that (1) a dividend tax penalty is incorporated into the return on a firm's common stock and (2) both a firm's dividend policy and its ownership structure impact the size of the dividend tax penalty. Prior related research, which uses the dividend yield to capture both the level of tax-disadvantaged dividend income and whether the marginal investor is more likely to be a low-tax or a high-tax investor, documents an increasing and concave relationship between a firm's dividend yield and its common stock return (e.g., Elton and Gruber, 1970; Litzenberger and Ramaswamy (1979); Litzenberger and Ramaswamy (1980) and Litzenberger and Ramaswamy (1982)). This finding is consistent with the notion that dividend taxes are capitalized into the return on a firm's common stock if it is ex-ante assumed that dividend tax clienteles exist such that investors in low tax brackets hold high-dividend-yield stocks and investors in high tax brackets hold low-dividend-yield stocks, resulting in a dividend tax penalty that is increasing at a decreasing rate with respect to the dividend yield. However, contrary to the notion of dividend tax clienteles, Jain (2000) finds that investors in low tax brackets tend to hold low-dividend-yield stocks and investors in high tax brackets tend to hold high-dividend-yield stocks. 1 This evidence suggests that it may not be appropriate to use a firm's dividend yield to capture whether the marginal investor is more likely to be a low-tax or a high-tax investor because this measure ex-ante assumes, perhaps incorrectly, that dividend tax clienteles exist. As in prior related research, our tests of whether the return on a firm's common stock incorporates a dividend tax penalty (i.e., our tests of the dividend tax capitalization hypothesis) use a firm's dividend yield to capture the level of tax-disadvantaged dividend income. although our evidence suggests that a higher level of institutional and corporate ownership reduces the dividend tax penalty if the dividend yield is held constant, it does not

provide an indication as to whether dividend tax clienteles are "properly" formed at a given dividend yield. The next section describes our data and our test variables. Section 3 discusses our test design and presents our major findings. Section 4 presents the results of several sensitivity analyses. The concluding section provides a summary of our findings.

Appendix B: Yes/No and open-ended survey questions

Please provide brief personal information:

1. How old are you?
2. Are you a male or a female?
3. What is your first language?
4. How many languages do you know?
5. How many semesters have you been at ISU or other US university?
6. What is your TOEFL score for admission to ISU?
7. Have you ever written a research article in your native language? If so, was it published? Where? What type of paper was it?
8. Have you ever written a research article in English? If so, was it published? Where? What type of paper was it?

Please read the following questions carefully. Write thoughtful and complete answers to these questions.

1. When working with IADE, has the program helped you focus on the moves? How? What exactly made you focus on them?
2. Did you learn the moves? What makes you say that?
3. What do you think helped you learn the moves the most?
4. Would you say that you learned the moves as a result of having focused on them? Why?
5. Would you say that you can transfer this knowledge to your writing? Why?
6. Has the program helped you notice anything about the moves/steps that you might not have paid much attention to before? How? What are these things?
7. Has the program helped you learn new words/expressions that signal particular moves/steps? How? Please provide examples of the words/expressions you learned for each move/step.
8. Do you think you improved your skill of writing a research article introduction? Why?
9. Did revising and resubmitting your drafts help you improve? Why?
10. Do you think it is helpful that IADE offers 'Help Options'? Why?

11. During the process of writing and revising your drafts, did you access any Help Options? Which Help Options did you use most often? Why?
12. Did you access the annotated corpus in the Help Options? Why?
13. When you were trying to revise the moves in your introduction, did you think of the meaning you wanted to express in those moves? Why?
14. When you were trying to revise the moves in your introduction, did you notice that what you meant in your move/s was reflected by a different color in the feedback? If so, what did you think and what did you do?
15. Do you think the feedback helped better express the meaning you had in mind? How?
16. Was the feedback appropriate for you? Why?
17. Was individualized feedback based on a comparison with the norms of writing in your field, motivating for you? Why?
18. What did you like about the feedback the most?
19. What did you like about the feedback the least?
20. What other kinds of feedback would be helpful to you?
21. Did IADE motivate you to improve your introduction? How?
22. Did you think a lot about how to improve? Why?
23. Were you excited to see improvement reflected in the feedback? Why?
24. Would you like to use this program again? When? Why?
25. Did the feedback influence your revision process or strategies in any way? How?

Appendix C. Likert-scale survey questions

1. What is your level of English writing skills?
   Excellent very good good fair poor

2. Did you focus on the moves?
   a lot somewhat a little not at all

3. How well did you learn the moves?
   Very well well a little not at all

4. Did you use IADE's Help Options?
   a lot somewhat a little not at all

5. Did your RA Introduction writing skills improve?
   a lot somewhat a little not at all

6. Did you notice any words/expressions that seemed to be characteristic of certain moves?
   a lot somewhat a little not at all

7. Did the feedback provided by IADE help you see the weaknesses in your drafts?
   a lot somewhat a little not at all

8. When you were revising, did you think about the meaning you wanted to express?
   a lot somewhat a little not at all

9. When you were revising, did you notice that you meant one move, but the feedback showed you the color of a different move?
   a lot somewhat a little not at all

10. Did the feedback help you think of what you meant and better express your ideas?

a lot somewhat a little not at all

11. Did you understand the feedback provided by the program?

completely somewhat a little not at all

12. The feedback provided by the program has a number of inherent characteristics. These are:
    - *Explicit* – the feedback is demonstrated clearly and fully
    - *Individual specific* – the feedback is provided to the student individually based on his/her input
    - *Metalinguisitc* – the feedback is provided in definitional terms (i.e., intro_m1, intro_m2, intro_m3)
    - *Color-coded* – the feedback about the actual draft structure is provided in colors: blue for Move 1, red for Move 2, and green for Move 3
    - *Negative* – the feedback points to drawbacks in the discourse development of the draft (e.g., the feedback may say: 'below average')
    - *Intelligent* – the feedback is generated automatically by a natural language processing-based system trained on a 1000 annotated RA introductions representative of 50 disciplines
    - *Output-focused* – the feedback targets your actual written production
    - *Iterative* – the feedback is provided as often as requested
    - *Short* – the feedback is provided concisely in that it briefly presents the descriptive percentages representing the use of Moves

How helpful did you find these characteristics of the feedback?

	Very helpful	Somewhat helpful	Of little help	Not helpful
explicit				
individual specific				
metalinguistic				
color-coded				
negative				
intelligent				
output-focused				
iterative				
short				

13. Was the feedback on discourse moves appropriate for you?

    very somewhat a little not at all

14. Did the program motivate you to improve your introduction draft?

    a lot somewhat a little not at all

15. Did you get enough writing practice with the program?

    a lot somewhat a little not at all

16. Did you like using IADE?

    much help some help little help no help

17. Did you get frustrated when using the program?

    a lot somewhat a little not at all

18. Did revising your draft with IADE have a positive impact on your skills of writing academic introductions?

    very somewhat a little not at all

19. Would you want to use a similar program to practice writing other research article sections?

    a lot somewhat a little not at all

Appendix D: Think-aloud protocol instructions

An essential data-gathering method in this research study is the think-aloud protocol. Think aloud protocols involve participants thinking aloud while they are performing a set of specified tasks.

The task for you, the participants, in this study is to interact with the IADE program by submitting your drafts of a RA introduction and revising your drafts based on the feedback returned by the program in order to improve the final product.

Please be aware that:

- It is not you who is investigated; it is the program that is being tested and evaluated. Any difficulties that you might be experiencing are not your fault.
- You can stop the task at any time if you become uncomfortable.
- You can use class materials if you need to access explanations of certain moves/steps
- You can use Callisto if you need to access examples of certain moves/steps
- You will not be told when you have completed the task; you must determine this on your own.

Please do the following when interacting with the program:

- Say whatever you are looking at
- Say whatever you are thinking
- Say whatever you are feeling
- Say whatever you are doing to go about the writing task

This will enable the researcher to see first-hand the process of task completion rather than only its final product.

The session will be audio and video taped <u>unobtrusively</u> (with screen capturing software) so that developers can go back and refer to what participants did and how they reacted. The purpose of this method is to make explicit what is implicitly present in the process of task completion.

References

Almond, R.G., Steinberg, L.S., & Mislevy, R.J. (2002). A four-process architecture for assessment delivery, with connections to assessment design. *Journal of Technology, Learning, and Assessment*, *1*(5), 1-63.

Anderson, J. R. (1982). *Acquisition of cognitive skill*. *Psychological Review*, *89,* 369-406.

Aston, G. (2002). The learner as corpus designer. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 9–25). Amsterdam: Rodopi.

Attali, Y. (2004). Exploring the feedback and revision features of the *Criterion* service. Paper presented at the *National Council on Measurement in Education Annual Meeting*, San Diego, CA. Retrieved on November 12, 2007 at http://www.ets.org/Media/Research/pdf/erater_NCME_2004_Attali_B.pdf

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater v.2. *Journal of Technology, Learning, and Assessment, 4*(3), 1-30.

Bates, L., Lane, J., & Lange, E. (1993). *Writing clearly: Responding to ESL compositions*. Boston: Heinle & Heinle.

Berkenkotter, D., Huckin, T.N., & Ackerman, J. (1991). Social context and socially constructed texts: The initiation of a graduate student into a writing research community. In C. Bazerman & J. Paradis (Eds.), *Textual Dynamics of the Professions: Historical and Contemporary Studies of Writing in Professional Communities* (pp. 191-215). Madison: University of Wisconsin Press.

Bernardini, S. (2000). Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 207–223). Frankfurt: Peter Lang.

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 165–182). Amsterdam: Rodopi.

Bhatia, V.K. (2002). A generic view of academic discourse. In J. Flowerdew (Ed.), *Academic Discourse* (pp. 21-39). Harlow: Longman.

Blyth, C., & Davis, J. (2007). Using formative evaluation in the development of learner-centered materials*, CALICO Journal, 25* (1), 48-68.

Braine, G. (2002). Academic literacy and the nonnative speaker graduate student. *Journal of English for Academic Purposes*, *1*, 59–68.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school* (expanded ed.), Washington, DC: National Academies Press.

Brown, S., & Knight, P. (1994). *Assessing Learners in Higher Education.* Kogan Page, London.

Buchanan, T. (1998). Using the World Wide Web for formative assessment. *Journal of Educational Technology Systems, 27*(1), 71–79.

Buchanan, T. (2000). Potential of the Internet for personality research. In M.H. Birnbaum (Ed.) *Psychological experiments on the Internet.* (pp. 121–140). San Diego: Academic Press.

Burstein, J. (2003). The e-rater text registered scoring engine: Automated essay scoring with natural language processing. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 113–121). Mahwah, NJ: Lawrence Erlbaum.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The *Criterion* Online Writing Evaluation service. *AI Magazine, 25*(3), 27-36.

Burstein, C., Marcu, D., & Knight, K. (2003). Finding the WRITE stuff: automatic identification of discourse structure in student essays. *IEEE Intelligent Systems*, *18*(1), 32–39.

Button, S. B., Mathieu, J. E., & Zajac, D. M. (1996). Goal orientation in organizational research: A conceptual and empirical foundation. *Organizational Behavior and Human Decision Processes, 67*, 26–48.

Byrne, D. (1986). *Teaching Oral English* (2nd ed.). Harlow, UK: Longman.

Cadman, K. (1997). Thesis writing for international students: A question of identity? *English for Specific Purposes*, *16*, 3–14.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Cargill, M., Cadman, K., & McGowan, U. (2001). Postgraduate writing: Using intersecting genres in a collaborative, content-based program. In I. Leki (Ed.), *Academic Writing Programs* (pp. 85-96). Alexandria, VA: TESOL.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, *15*, 357–366.

Casanave, C., & Hubbard, P. (1992). The writing assignments and writing problems of doctoral students: Faculty perceptions, pedagogical issues, and needed research. *English for Specific Purposes*, *11*, 33–49.

Chapelle, C. A. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning & Technology, 2*(1), 22-34.

Chapelle, C. A. (2001). *Computer applications in second language acquisition.* Cambridge: Cambridge University Press.

Chapelle, C. A. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology*. Amsterdam: John Benjamins.

Chapelle, C. A. (2007). Technology and second language acquisition. *Annual Review of Applied Linguistics*, *27*, 98–114.

Chapelle, C. A. (2001). *Computer applications in second language acquisition.* Cambridge: Cambridge University Press.

Chapelle, C. A., Jamieson, J., & Park, Y. (1996). Second language classroom research traditions: How does CALL fit? In M. Pennington (Ed.), *The power of CALL* (pp. 33-53). Houston, TX: Athelstan.

Chen, C. F, & Cheng, W. Y., (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning and Technology, 12*(2), 94-112.

Chen, L., & Tokuda, N. (2003). A new template-template-enhanced ICALL system for a second language composition course. *CALICO Journal*, *20*(3), 561-578.

Chen, L., Tokuda, N., & Xiao, D. (2002). A POST parser-based learner model for template-based ICALL for Japanese-English writing skills. *Computer Assisted Language Learning*, *15*(4), 357-372.

Cheng, W., Warren, M., & Xu, X. (2003). The language learner as language researcher: Putting corpus linguistics on the timetable. *System, 31*(2), 173–186.

Cheville, J. (2004). Automated scoring technologies and the rising influence of error. *English Journal*, *93*(4), 47–52.

Chung, G., & Baker, E. (2003). Issues in the reliability and validity of automated scoring of constructed responses. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 23-40). Mahwah, NJ: Lawrence Associates.

Cook, V. J. (1991). *Second language learning and language teaching*. London: Edward Arnold.

Cooley, L., & Lewkowicz, J. (1995). The writing needs of graduate students at the University of Hong Kong: A project report. *Hong Kong Papers in Linguistics and Language Teaching*, *18*, 121–123.

Cooley, L., & Lewkowicz, J. (1997). Developing awareness of the rhetorical and linguistic conventions of writing a thesis in English: Addressing the needs of ESL/EFL

postgraduate students. In A. Duszak (Ed.), *Culture and styles of academic discourse* (pp. 113–140). Berlin: Mouton de Gruyter.

Conference on College Composition and Communication (2006). *Writing assessment: A position statement*. Retrieved October 10, 2008, from http://www.ncte.org/cccc/resources/positions/123784.htm

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In *Proceedings of ACM CHI 2001 conference on human factors in computing systems* (pp. 245–252). New York: ACM Press.

Cortes, V. (2007). Exploring genre and corpora in the English for academic writing class. *The ORTESOL Journal*, *25*, 8-14.

Creswell, J. W. (2003). Research Design: Qualitative, quantitative, and mixed methods approaches. London: SAGE Publishers.

Pendar, N., & Cotos, E. (2008). Automatic identification of discourse moves in scientific article introductions. In *The Proceedings of The 3rd workshop on innovative Use of NLP for Building EducationalApplications* (pp. 62-70). Columbus, Ohio.

DeKeiser, R. (2007). Skill acquisition theory. In J. Williams & B. VanPatten (Eds.), *Theories in Second Language Acquisition: An introduction* (pp. 97-113). Mahwah, NJ: Erlbaum.

Dempsey, J. V., Discoll, M. P., & Swindell L. K. (1993). Text-based feedback. In J. V. Dempsey & G. C. Sales (Eds.), *Interactive instruction and feedback* (pp. 21-53). Englewood Cliffs, N.J.: Educational Technology Publications.

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning, and Assessment*, *5*(1), 4-35.

Dong, Y. (1998). Non-native graduate students' thesis/dissertation writing in science: Self-reports by students and their advisors from two US institutions. *English for Specific Purposes, 17*, 369–390.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040–1048.

Dweck, C. S., & Leggett, E. L. (1988). *A social-cognitive approach to motivation and personality*. Psychological Review, *95*(2), 256-273.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of Educational Measurement* (5th ed.). Prentice Hall, Inc. Englewood Cliffs, NJ.

Elliot, S. (2003). IntelliMetric: From here to validity. In M. D. Shermis & J.C. Burstein, *Automated essay scoring: A cross-disciplinary perspective* (pp. 71-86). Mahwah, NJ: Lawrence Erlbaum Associates.

Elliot, S., & Mikulas, C. (2004). *The impact of MY Access!™ use on student writing performance: A technology overview and four studies.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Ellis, R. (1999). *Learning a second language through interaction.* Amsterdam: John Benjamins.

Ericsson, P. (2006). The meaning of meaning: Is a paragraph more than an equation? In P. F. Ericsson & R. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 28-37). Logan, UT: Utah State University Press.

Ericsson, P., & Haswell, R. (Eds.) (2006). *Machine scoring of student essays: Truth and consequences.* Logan, UT: Utah State University Press.

Ericsson, K., & Simon, H. (1987). Verbal reports on thinking. In C. Faerch & G. Kasper (Eds.). *Introspection in Second Language Research* (pp. 24–54). Clevedon, Avon: Multilingual Matters.

Fisher, S. L., & Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology, 51*, 397–420.

Farr, J. L., Hofmann, D. A., & Ringenbach, K. L. (1993). Goal orientation and action control theory: Implications for industrial and organizational psychology. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 193–232). New York: Wiley.

Ferris, D. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29,* 33-53.

Ferris, D. (2003). *Response to student writing: Implications for second language students.* Mahwah, NJ: Lawrence Erlbaum.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). Cambridge University Press.

Fitts, P., & Posner, M. (1967). *Human performance*. Belmont, CA: Brooks/Cole.

Flowerdew, L. (1998). Concordancing on an expert and learner corpus for ESP. *CALL Journal, 8* (3), 3–7.

Foltz, P., Laham, D., & Landauer, T. (1999). The Intelligent Essay Assessor: Applications to Educational Technology. *Interactive Multimedia Electronic Journal of Computer-Enhanced Leaning, 1*(2). Retrieved on December 8, 2008 at http://imej.wfu.edu/articles/1999/2/04/index.asp

Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology, 83*, 218–233.

Frodesen, J. (1995). Negotiating the syllabus: a learning-centered, interactive approach to ESL graduate writing course design. In D. Belcher & G. Braine (Eds.) *Academic Writing in a Second Language: Essays on Research and Pedagogy*. (pp. 331–350). Norwood, NJ: Ablex Publishing.

Frost, A. (1999). Supervision of NESB postgraduate students in science-based disciplines. In Y. Ryan & O. Zuber-Skerritt (Eds.), *Supervising postgraduates from non-speaking backgrounds* (pp. 101–109). Birmingham, UK: Open University Press.

Garner, M., & Borg, E. (2005). An ecological perspective on content-based instruction. *Journal of English for Academic Purposes, 4*, 119–134

Garrett, N. (1987). A Psycholinguistic Perspective on Grammar and CALL. In Wm. Flint Smith (Ed.) *Modern Media in Foreign Language Education: Theory and Implementation*. (pp. 169-96) Lincolnwood, IL: National Textbook.

Garrett, N. (1991). Technology in the service of language learning: Trends and issues. *The Modern Language Journal, 75*(1), 74-101.

Gass, S., & Mackey, A. (2007). Input, interaction, and output in Second Language Acquisition. In J. Williams & B. VanPatten (Eds.), *Theories in Second Language Acquisition: An introduction*. (pp. 175-199). Mahwah, NJ: Erlbaum.

Ghadessy, M., Henry, A., & Roseberry, R. (Eds.). (2001). *Small corpus studies and ELT: Theory and practice*. Amsterdam/Philadelphia: John Benjamins.

Halliday, M.A.K. (1971). Language in a social perspective. *The Context of Language: Educational Review*, *23*(3), 165-188. University of Birmingham.

Halliday, M.A.K. (1982). How is a text like a clause? S. Allén (Ed.), *Text Processing.* Stockholm: Almqvist & Wiksell.

Halliday, M. A. K. (1985). Systemic background. In J. D. Benson & W. S. Greaves, (Eds.) *Systemic perspectives on discourse, Volume 1, volume XV in the series Advances in discourse processes* (pp. 1- 15). Norwood, NJ: Ablex Publishing Corporation.

Halliday M.A.K. (1989a). Functions of language. In Halliday, M.A.K. & R. Hasan, *Language, context, and text* (pp. 15-26). Oxford University Press, Oxford, UK.

Halliday M.A.K. (1989b). Register variation. In Halliday, M.A.K. & R. Hasan, *Language, context, and text* (pp. 29-43). Oxford University Press, Oxford, UK.

Halliday M.A.K. (2007). *Language and society*. J. Webster, (Ed.). London, UK.

Halliday, M. A. K., & J. R., Martin. (1993). *Writing science: Literacy and discursive power*. London: Falmer.

Halliday, M.A.K., & R. Hasan. (1976). *Cohesion in English.* Longman.

Hasan, R. (1978). Text in the systemic-functional model. In W. U. Dressler (Ed.), *Current trends in textlinguistics* (pp. 228-246). duGruyter, Berlin & New York.

Hasan, R. (1989). The structure of a text. In M.A.K. Halliday & R. Hasan, *Language, context, and text* (pp. 52-68). Oxford University Press, Oxford, UK.

Holland, M., Maisano, R., Alderks, C., & Martin, J. (1993). Parsers in Tutors: What Are They Good For? *CALICO Journal, 11*(1), 28-46.

Heift, T. (2001). Error-specific and individualized feedback in a web-based language tutoring system: Do they read It? *ReCALL, 13*(2), 129-142.

Heift, T. (2002). Learner control and error correction in ICALL: Browsers, peekers and adamants. *CALICO Journal, 19*(3), 295-313.

Heift, T. (2003). Multiple Learner Errors and Meaningful Feedback: A Challenge for ICALL Systems. *CALICO, 20*(3), 533-549.

Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*, *16*(2), 416-431.

Heift, T. (2005). Inspectable learner reports for web-based language learning. *ReCALL, 17*(1), 32-46.

Heift, T. (2006). Context-sensitive Help in CALL. *Computer Assisted Language Learning, 19*(2), 243-259.

Heift, T. (2008). Corrective feedback in CALL: A longitudinal study of learner uptake. Paper presented at the AAAL 2008 Conference, Washington DC.

Heift, T., & Nicholson, D. (2001). Web Delivery of Adaptive and Interactive Language Tutoring. *International Journal of Artificial Intelligence in Education, 12*(4), 310-325.

Heift, T., & Rimrott, A. (2008). Learner Responses to Corrective Feedback for Spelling Errors in CALL. *System*, *36*(2) 1-18.

Heift, T., Toole, J., McFetridge, P., Popwich, & F., Tsiplakou, S. (2000). Learning Greek with an Intelligent and Adaptive Hypermedia System. *IMEJ of Computer-Enhanced Learning*, *2*(2), October 2000 (http://imej.wfu.edu).

Hoska, D. M. (1993). Motivating learners through CBI feedback: Developing a positive learner perspective. In V. Dempsey & G. C. Sales (Eds.). *Interactive instruction and*

feedback (pp. 105-132). Englewood Cliffs, N.J.: Educational Technology Publications.

Hubbard, P. (1988). An integrated framework for CALL courseware evaluation. *CALICO Journal, 6*(2), 51-72.

Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing*, *7*(3), 255-286.

Hyland, K. (2000). *Disciplinary discourses: social interactions in academic writing*. Harlow, Essex: Longman.

Hyland, K. (2002). Specificity revisited: How far should we go now? *English for Specific Purposes*, *21*(4), 385–395.

Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.

Hyland, K., & Hyland, F. (Eds.). (2006). *Feedback in Second Language Writing: Contexts and issues.* New York: Cambridge University Press.

James, K. (1984). The writing of theses by speakers of English as a foreign language: A case study. In R. Williams, J. Swales, & J. Kirkman (Eds.), *Common ground: Shared interests in ESP and communication studies, ELT documents* (pp. 99–113). Oxford: Pergamon Press.

Jamieson, J., Chapelle, C.A., & Preiss, S. (2004). Putting principles into practice. *ReCALL, 16*(2), 396–415.

Jamieson, J., Chapelle, C.A., & Preiss, S. (2005). CALL Evaluation by developers, a teacher, and students, *CALICO Journal,* 23(1), 93-138.

Jenkins, S., Jordan, M., & Weiland, P. (1993). The role of writing in graduate engineering education: A survey of faculty beliefs and practices. *English for Specific Purposes, 12*, 51–67.

Juozulynas, V. (1994). Errors in the Compositions of Second-Year German Students: An Empirical Study for Parse-Based ICALI. *CALICO Journal, 12*(1), 5-17.

Keith, T. (2003). Validity and automated essay scoring systems. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective*. (pp. 147-167). Mahwah, NJ: Lawrence Associates.

Kettemann, B., & G. Marko (Ed). (2002). *Teaching and learning by doing corpus analysis*. Amsterdam: Rodopi.

Knight, N. (1999). Responsibilities and limits in the supervision of NESB research students in the social sciences and humanities. In Y. Ryan, & O. Zuber-Skerritt (Eds.), *Supervising postgraduates from non-English speaking backgrounds* (pp. 93–100). Buckingham, UK: Open University Press.

Krashen, S. (1982). *Principles and practices in second language acquisition*. Oxford: Pergamon.

Krashen, S. (1984). *Writing: Research, theory, and application*. Oxford: Pergamon Press.

Krashen S. (1985) *The input hypothesis: issues and implications*, London: Longman.

Kulhavy, R. W., & Stock, W. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279–308.

Kushner, S. (1997). Tackling the needs of foreign academic writers: A case study. *IEEE Transactions on Professional Communication*, *40*, 20–25.

Labrie, G., & Singh, L. (1991). Parsing, error diagnostics, and instruction in a French Tutor. *CALICO Journal, 9*(1), 9-25.

Lalande, J. F. (1982). Reducing composition errors: An experiment. *The Modern Language Journal*, *66*, 140-149.

Landauer, T., Laham, D., and Foltz, P. (2003). Automated Scoring and Annotation of Essays with the Intelligent Essay Assessor. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective*. (pp. 87-112). Mahwah, NJ: Lawrence Erlbaum Associates.

Leah Rock, J. (2007). The impact of short-term use of *CriterionSM* on writing skills in ninth grade. Research report. Educational Testing Service, Princeton, NJ. Retrieved on December 8, 2008 from http://www.ets.org/Media/Research/pdf/RR-07-07.pdf

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*, 56–75.

Levin, L., Evans, D., & Gates, D. (1991). The ALICE system: A workbench for learning and using language. *CALICO, Journal 9*(1), 27-56.

Levis, J., & Muller-Levis, G. (2003). A Project-based approach to teaching research writing to nonnative writers. *IEEE Transactions on Professional Communication*, *46*(3), 210-220.

Liou, H. C. (1991). Development of an English Grammar Checker: A progress report. *CALICO Journal, 9*(1), 57-70.

Long, M. (1984). Process and product in ESL programme evaluation. *TESOL Quarterly 18(3)*, 409–25.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. K. Bhatia (Eds.), *Handbook of language acquisition: Vol. 2. Second language acquisition* (pp. 413-468). San Diego, CA: Academic Press.

Loritz, D. (1992). Generalized transition network parsing for language study: The GPARS system for English, Russian, Japanese, and Chinese. *CALICO Journal, 10*(1), 5-22.

Lyster, R. (1998). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, *48*, 183–218.

Ma, Q. (2008). Empirical CALL evaluation: The relationship between learning process and learning outcome, *CALICO Journal, 26*(1), 108-122.

Mason, B. J., & Bruning, R. (2001). Providing feedback in computer-based instruction: What the research tells us. Retrieved December, 2, 2008, from http://dwb4.unl.edu/dwb/Research/MB/MasonBruning.html

Martin, J. R. (1985). Process and text. In J.D. Benson & W.S. Greaves (Eds.), *Systemic perspectives on discourse* (Vol. 1, pp. 248-274). Norwood, NJ: Ablex Publishing.

Martin, J.R. (1992). *English text: systems and structure*. Benjamins, Amsterdam.

Matsuda, P. K. (2003). Second language writing in the twentieth century: A situated historical perspective. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 15-34). New York: Cambridge University Press.

McMillan, J. H. (2007). *Formative Classroom Assessment: Theory into practice*. NY: Teachers College Press.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*, 241-256.

Mislevy, R., Almond, R., and Lukas, J. (2003). A brief introduction to Evidence Centered Design. Report 632. Retrieved August 27, 2008 from http://www.education.umd.edu/EDMS/mislevy/papers/

Mitchell, T., Russel, T., Broomhead, P., and Aldridge, N. (2002). Towards Robust Computerised Marking of Free-Text Responses. In: *Proceedings of the 6th CAA Conference*, Loughborough: Loughborough University. Retrieved on July 31, 2007 at https://dspace.lboro.ac.uk/dspace/bitstream/2134/1884/1/Mitchell_t1.pdf.

Mory, E. H. (2004). Feedback research review. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745–783). Mahwah, NJ: Erlbaum Associates.

Nagata, N. (1993). Intelligent computer feedback for second language instruction, *The Modern Language Journal, 77*(3), 330-339.

Nagata, N. (1995). An effective application of natural language processing in second language Instruction. *CALICO Journal, 13*(1), 47–67.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Ed.), *Instructional design for multimedia learning* (pp. 181–195). Munster, New York: Waxmann.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nwongu, K. N. (1990). *Discourse variation in medical texts: Schema, theme and cohesion in professional and journalistic accounts.* University of Nittingham: Monographs in Systemic Linguistics, 2.

Page, E. (2003). Project Essay Grade. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp.43-54). Mahwah, NJ:Lawrence Associates.

Parry, S. (1998). Disciplinary discourse in doctoral education. *Higher Education, 36*, 273–299.

Pearson Education, Inc. (2007). General overview of WriteToLearn™ and its components. Retrieved on December 8, 2008 from http://www.knowledge-technologies.com/papers/WTLOverview-101507.pdf

Pica, T. (1998). Second language learning through interaction: Multiple perspectives. In V. Regan (Eds.), *Contemporary approaches to second language acquisition* (pp. 1-31). Dublin: University of Dublin Press.

Pridemore, D. R., & Klein, J. D. (1995). Control of practice and level of feedback in computer-based instruction. *Contemporary Educational Psychology, 20*, 444–450.

Reuer, V. (2003). Error recognition and feedback with Lexical Functional Grammar. *CALICO Journal*, *20*(3), 497-512.

Riazi, A. (1997). Acquiring disciplinary literacy: A social-cognitive analysis of text production and learning among Iranian graduate students of education. *Journal of Second Language Writing*, *6*, 105–137.

Rosa, E., & Leow, R. (2004). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics*, *25*, 269–292.

Rudner, L., & Liang, T. (2002). Automated essay scoring using Bayes' theorem. *The Journal of Technology, Learning and Assessment*, *1*(2), 3-21.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119-144.

Sales, G. C. (1993). Adapted and adaptive feedback in technology-based instruction. In V. Dempsey & G. C. Sales (Eds.), *Interactive instruction and feedback* (pp. 159–175). Englewood Cliffs, N.J.: Educational Technology Publications.

Sanders, R. (1991). Error Analysis in Purely Syntactic Parsing of Free Input: The Example of German. *CALICO Journal, 9*(1), 72-89.

San Miguel, C., & Nelson, C. (2007). Key writing challenges of practice-based doctorates. *Journal of English for Academic Purposes*, *6*, 71–86.

Shaalan, K. (2005). An intelligent computer assisted language learning system for Arabic learners. *Computer Assisted Language Learning*, *18*,(1-2), 81-108.

Shermis, M., & Burstein, J. (2003). Introduction. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective*. (pp. xiii-xvi). Mahwah, NJ: Lawrence Erlbaum.

Shermis, M., Burstein, J., & Bliss, L. (2004). *The impact of automated essay scoring on high stakes writing assessments.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, *13*, 298-317.

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153-189.

Si-Qing, C., & Luomai, X. (1990). Grammar-debugger: A parser for Chinese learners. *CALICO Journal*, *8*(2), 63–75.

Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. *Research in the Teaching of English, 31,* 91-119.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp.125-144). Oxford: Oxford University press.

Swales, J. (1981). *Aspects of articles introductions.* Aston ESP Reports No. 1. The University of Aston in Birmingham.

Swales, J.M. (1990). *Genre analysis: English in academic and research settings.* Cambridge: Cambridge University Press.

Swales, J.M. (2004). *Research genres: Exploration and applications.* Cambridge: Cambridge University Press.

Thompson, P. (1999). Exploring the contexts of writing: Interviews with Ph D supervisors. In P. Thompson (Ed.), *Issues in EAP writing research and instruction* (pp. 37–54). Reading, PA: Centre for Applied Language Studies, University of Reading.

Thompson, P., & Tribble, C. (2001). Looking at citations: Using corpora in English for Academic Purposes. *Language Learning and Technology, 5*, 91–105.

Tokuda, N., & Chen, L. (2004). A new KE-free online ICALL system featuring error contingent feedback. *Computer Assisted Language Learning*, *17*(2), 177-201.

Tokuda, N., & Chen, L. (2001). An online tutoring system for language translation. *IEEE Multimedia, 8*(3), 46-55.

Toole, J., & Heift, T. (2002). The Tutor Assistant: An authoring system for a Web-based intelligent language tutor. *Computer Assisted Language Learning*, *15*(4), 373-86.

Torrance, H., & Pryor, J. (1998). *Investigating Formative Assessment. Teaching, Learning and Assessment in the Classroom*. Buckingham, Open University Press.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, *46*, 327-369.

van der Linden, E. (1993). Does feedback enhance computer-assisted language learning. *Computers & Education, 21*(1-2), 61-65.

Vann, R., & Myers, C. (2001). Academic ESL options in a large research university. In I. Leki (Ed.), *In Academic Writing Programs* (pp. 73-84). Alexandria, VA: TESOL.

Vantage Learning. (2007). *MY Access!* Efficacy Report**.** Newtown, PA: Vantage Learning. Retrieved on December 6, 2008, from http://www.vantagelearning.com/school/research/myaccess.html

Ventola, E. (1982). *Contrasting schematic structures in service encounters*. University of Sydney: Mimeo.

Vygotsky, L. S. (1987). *The collected works of L.S. Vygotsky, Vol.1: Problems of general psychology*, tr. By Norris Minnik. New York: Plenum.

Ware, P. (2005). Automated writing evaluation as a pedagogical tool for writing assessment. In A. Pandian, G. Chakravarthy, P. Kell, & S. Kaur (Eds.), *Strategies and practices for improving learning literacy* (pp. 174-184). Selangor, Malaysia: University Putra Malaysia Press.

Warschauer, M., & Grimes, G. (2008). Automated Writing Assessment in the Classroom. *Pedagogies: An International Journal,* 3, 22–36.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research, 10*(2), 1–24.

Whithaus, C. (2006). Always already: Automated essay scoring and grammar checkers in college writing courses. In P. F. Ericsson & R. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 166-176). Logan, UT: Utah State University Press.

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, *22*, 537-548.

Williamson, M. (2004). Validity of automated scoring: Prologue for a continuing discussion of machine scoring student writing. *Journal of Writing Assessment, 1*(2), 85–104.

Yang, N. (2004). Using MyAccess in EFL writing. In *The proceedings of 2004 International Conference and Workshop on TEFL & Applied Linguistics* (pp. 550-564). Taipei, Taiwan: Ming Chuan University.

Yang, J.C., & Akahori, K. (1998). Error analysis in Japanese writing and its implementation in a computer assisted language learning system on the world wide web. *CALICO Journal*, *15*(1-3), 47-66.

Yang, J.C., & Akahori, K. (1999). An evaluation of Japanese CALL systems on the WWW comparing a freely input approach with multiple selection. *Computer Assisted Language Learning*, *12*(1), 59-79.

Yazdani, M. (1991). The LINGER project: An artificial intelligence approach to second-language tutoring. *Computer Assisted Language Learning, 4*(2), 107-116.